MODERN CELESTIAL MECHANICS

Aspects of Solar System Dynamics

Alessandro Morbidelli

November 4, 2011

Preface

In the last 20 years, Celestial Mechanics has achieved spectacular results on the structure and evolution of our Solar System. The discovery of the chaotic dynamics of the planets, the identification of the main transport routes by which asteroids escape from the main belt and reach Earth-threatening orbits, and the understanding of the dynamical mechanisms at the origin of the internal heating of the Galilean satellites, are just a few examples of the results that have made Celestial Mechanics a respectable branch of Solar System science. As a consequence, worldwide congresses on planetology, such as the annual meetings of the Division for Planetary Sciences of the American Astronomical Society or those of the Asteroids Comets & Meteoroids series, always schedule long and well-attended sessions on dynamics.

This is what I call *Modern Celestial Mechanics*, although one should not forget that other facets of Celestial Mechanics, concentrated more on the rigorous mathematical study of toy models rather than on the realistic description of Solar System evolution, have also recently achieved innovative results.

Several books have been published on Celestial Mechanics, but none of them describe the recent results on Solar System dynamics, with a solid theoretical basis. They are either focused on the general theory of dynamical systems or limited to the fundamentals of Celestial Mechanics. The present book should fill this gap, in the hope of helping advanced students and young researchers, who are nowadays compelled to find their way through a vast literature of scientific papers without the aid of a guide book presenting the state of the art in a unified form. The goal is to take the reader to a point where he can start developing his own original contribution.

Modern Celestial Mechanics is intimately related to the theory of Hamiltonian systems. Most analytic studies make an essential use of Hamiltonian perturbation techniques; but also the correct interpretation of the results of numerical simulations often requires a good theoretical knowledge of Hamiltonian dynamics. Because Hamiltonian theory does not enter in the usual cultural baggage of people approaching Celestial Mechanics from the astronomical side, the first part of this book revisits its fundamental concepts. Without the pretence of being exhaustive, the first six chapters present what one should know of Hamiltonian theory to work at ease in Celestial Mechanics.

Details and technical mathematical proofs are skipped, while only the guiding ideas are given, with specific references for those who would like to enter deeper into the subject. After the first chapter on the basic concepts of Celestial and Hamiltonian Mechanics, Chapter 2 explains Hamiltonian perturbation theory based on Lie series. Chapters 3 and 4 illustrate the properties of invariant tori and resonances, respectively. Particular care is paid in Chapter 5 to discussing the numerical tools that are useful for the detection of chaos. Chapter 6 discusses the possible dynamical structures of Hamiltonian systems that result from the interaction of its resonances, and details how these structures can be identified with numerical explorations.

The second part of the book is devoted to the subject of Celestial Mechanics itself. This part is more technical, although care is taken to concentrate on the procedures and on the ideas, and not on the most technical details. Thus, the book should also be pleasant for reading experts in Hamiltonian theory, who are curious to know what is done in Modern Celestial Mechanics. In particular, this book does not go into detail on the techniques for practical computations (series expansion or close evaluation of the perturbation functions, numerical calculation of the action integrals, etc.). In fact, practical techniques evolve very fast, in parallel with the evolution of computer power, while the ideas and the conceptual approaches stay valid for a much longer time. Chapters 7 and 8 are respectively devoted to the secular motion of planets and of small bodies. The chaotic dynamics of the terrestrial planets, the theories for the computation of an asteroid's proper elements and the dynamics of secular resonances are among the issues treated in these chapters. Chapters 9-12, conversely, are devoted to the difficult subject of mean motion resonances. In particular, Chapter 9 outlines the structure of mean motion resonances, first in the framework of the simple planar restricted three-body problem, and then in more realistic models. Chapter 10 is on three-body resonances, whose importance has been recently understood both for small body dynamics and for planetary dynamics. Chapter 11 discusses the secular dynamics *inside* mean motion resonances, which in my opinion is one of the most complicated topics of Celestial Mechanics. Finally Chapter 12 investigates the global dynamical structure of the regions of the Solar System that are densely inhabited by small bodies, and discusses the fashionable subject of slow chaotic diffusion.

As one can see, the book is strongly focused on the dynamics of planets and of small bodies. This is mainly due to the limitation of my knowledge. Important problems related to the dynamics of natural satellites and of planetary rings (see the Introduction for an overview) are therefore not discussed. However, most of the concepts developed here apply also to satellite and ring

dynamics, so that I hope that this book can serve as an introduction also for people interested in these fields. Moreover, I have decided to exclude from this book the dynamics of bodies that have close encounters with the planets, because it does not fall into the category of quasi-integrable dynamics, and its relationship with the rest of the book would be very weak. In fact, the study of planet crossing dynamics is essentially numeric, the possibilities of analytic and semi-analytic theories being very limited.

The completion of this book would not have been possible without the encouragement of several colleagues, first of all M. Festou and C. Froeschlé. I'm in debt to the main revisors – M. Guzzo, D. Nesvorny and F. Thomas – for their constructive criticism, and to the reviewers – A. Giorgilli, V. Gurzadyan, J. Henrard and J. Laskar – for their appreciation and suggestions. At least seven readers had the courage to go through every line of this book!

I also wish to thank A. Cellino, S. Ferraz-Mello, J. Henrard, M. Holman, Z. Knežević, J. Laskar, E. Lega, A. Lemaitre, C. Murray, N. Murray and D. Nesvorný for providing the original printouts of their figures.

I dedicate this book to the memory of my friends and colleagues Paolo Farinella, Fabio Migliorini, and Michèle Moons. Life did not give them a second chance, and I'm missing their affection and their advice.

La Turbie (France), June 1, 2000.

Contents

PREFACE **i**

INTRODUCTION **1**

1 ELEMENTARY CELESTIAL AND HAMILTONIAN MECHANICS **11**

1.1 Equations of motion 11
1.2 Orbital elements 12
1.3 Perturbations of the two-body problem 16
1.4 Hamiltonian systems and the two-body problem 18
1.5 Perturbations in Hamiltonian form 20
1.6 Canonical transformations 22
1.7 Properties of Hamiltonian flow 25
1.8 Integrable Hamiltonians 26
1.9 Action–angle variables 28
1.9.1 Delaunay variables 29
1.9.2 Hamilton equations in Delaunay variables for the restricted and the planetary problem 33
1.9.3 D'Alembert rules 35
1.10 Integrable dynamics 36

2 QUASI-INTEGRABLE HAMILTONIAN SYSTEMS **39**

2.1 Introduction to perturbation theory 39
2.2 Lie series approach 40
2.3 The small divisors problem 42
2.3.1 Normal forms 45
2.4 Beyond the first order 46
2.4.1 Example of computation of the optimal order of the normal form 48

2.4.2 Generation of high-order harmonics by the normalization process 49
2.5 Averaging over the mean motions 51
2.5.1 Secular normal form 53
2.5.2 Mean motion resonant normal form 57

3 KAM TORI 59
3.1 Kolmogorov's theorem 59
3.1.1 Sketch of proof of Kolmogorov's theorem 60
3.2 Properties of KAM tori 63
3.3 Numerical examples 67

4 SINGLE RESONANCE DYNAMICS 71
4.1 The integrable approximation 71
4.2 Resonant action–angle variables 74
4.3 Perturbed resonant dynamics 78
4.3.1 Size of the remainder 78
4.3.2 Resonant invariant tori 79
4.3.3 Splitting of separatrices 81
4.3.4 Size of the chaotic region 84

5 NUMERICAL TOOLS FOR THE DETECTION OF CHAOS 89
5.1 Monitoring the time evolution in phase space 89
5.2 Lyapunov exponents 92
5.2.1 Numerical computation of the MLE 94
5.3 Frequency analysis 98
5.3.1 Numerical determination of the frequencies 102
5.4 Surrogates 103
5.4.1 Fast Lyapunov indicator 103
5.4.2 Helicity and twist angles 104
5.4.3 Mean exponential growth factor of nearby orbits 105
5.4.4 Mean, maximal and minimal values of the actions . . . 106

6 INTERACTIONS AMONG RESONANCES 107
6.1 Two degrees of freedom 107
6.1.1 Heteroclinic intersections 111
6.2 More than two degrees of freedom 114
6.2.1 The Nekhoroshev theorem 114
6.2.2 Nekhoroshev structure 119
6.2.3 Superexponential stability of KAM tori 122
6.3 Exploring the dynamical structure of a given system 122

7 SECULAR DYNAMICS OF THE PLANETS 127
7.1 Lagrange–Laplace solution 127
7.2 Higher-order solutions 131
7.3 Chaotic secular motion of the planets 135
7.4 Spin axes dynamics 140

8 SECULAR DYNAMICS OF SMALL BODIES 149
8.1 The linear integrable approximation 149
8.2 The Kozai integrable approximation 154
8.2.1 Kozai dynamics inside the orbit of the main perturber . 156
8.2.2 Kozai dynamics outside the orbit of the main perturber 160
8.2.3 Action–angle variables for the Kozai Hamiltonian 166
8.3 Proper elements 167
8.3.1 Asteroid families 172
8.4 Secular resonances 174
8.4.1 Secular resonant dynamics 179
8.4.2 The anomalous case of the ν_6 resonance 187

9 MEAN MOTION RESONANCES 195
9.1 A simple integrable approximation 195
9.1.1 Phase protection from planetary collisions 206
9.1.2 The case of the 1/1 resonance 208
9.2 Mean motion resonance overlapping 210
9.2.1 Threshold for overlapping in the vicinity of the planet . 216
9.2.2 Overlapping of resonances with different planets 217
9.3 Resonant multiplets 220
9.4 The modulated pendulum approximation 224

10 THREE-BODY RESONANCES 231
10.1 Origin of the resonant perturbation terms 231
10.1.1 The direct effect 232
10.1.2 The indirect effect 234
10.1.3 Inclusion of both direct and indirect effects in the asteroid problem 239
10.2 Three-body resonant multiplets 240
10.3 The asteroid and Kuiper belts 247
10.4 Chaotic dynamics of the giant planets 254

11 SECULAR DYNAMICS INSIDE MEAN MOTION RESONANCES 257
11.1 Successive elimination of harmonics 257

11.2 The mean motion resonant dynamical system 260
11.2.1 Secondary resonances 262
11.2.2 The Kozai dynamics 266
11.2.3 Perihelion secular resonances 271
11.2.4 Nodal secular resonances 276
11.2.5 Three-body resonances 277
11.3 The major resonances in the asteroid belt 280
11.3.1 The 3/1 resonance 280
11.3.2 The 2/1 resonance 286
11.3.3 The 3/2 resonance 292
11.4 The major resonances in the Kuiper belt 294
11.4.1 The 2/3 resonance 295
11.4.2 The 1/2 resonance 298
11.5 The 1/1 resonances 300

12 GLOBAL DYNAMICAL STRUCTURE OF THE BELTS OF SMALL BODIES 303
12.1 Detection of the chaotic zones 303
12.2 Chaotic diffusion and macroscopic instability 312
12.3 Analytic estimates of Lyapunov time and instability time 320
12.4 Do the KAM and Nekhoroshev theorems apply for small body dynamics? 327

REFERENCES 331

Introduction

OBJECTS OF THE SOLAR SYSTEM, PROBLEMS OF CELESTIAL MECHANICS

This chapter presents a general overview of the Solar System for nonexpert readers. The goal is to show that almost every object in the Solar System raises dynamical problems, which makes Celestial Mechanics a fundamental branch of Solar System science. Some of these problems will be treated in detail in the following chapters of the book.

The first attempts to predict the motion of the planets in the sky date back to the dawn of civilization. However, what is now properly called *Celestial Mechanics* was born with Newton's announcement (1687) of the law of universal gravitation. Since then, the computation of precise ephemerides has become essentially a mathematical challange. Lagrange and Laplace particularly distinguished themselves in this search at the end of the XVIIIth century. To obtain increasing accuracy, they designed the first step of the theory of secular motion of planetary orbits, and the latter also understood the importance of a quasi-resonance between the orbital periods of Jupiter and Saturn. At that time, every apparent anomaly in planetary motion seemed to be accountable by a sufficiently accurate theory based on Newtonian gravitation. This success forged Laplace's absolute faith in determinism, which then characterized the whole science of the XIXth century. The search for increasingly accurate analytic ephemerides has been continuing through the centuries up to the present time. But two centuries later than Laplace's work – after the discovery of a new phenomenon called *chaos*, the introduction of the first numerical methods for its detection and the construction of powerful computers – it was pointed out that the motion of the planets is in fact chaotic. Since then, a major goal of Celestial Mechanics has been to understand the reasons for this chaotic mo-

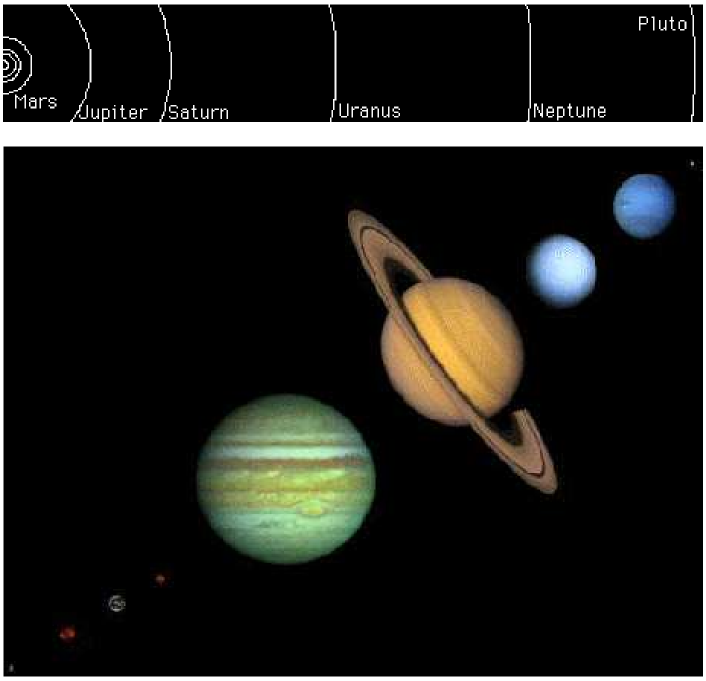


Figure 1: Top: the relative sizes of the orbits of the nine planets (the orbits of Mercury, Venus and Earth are not labeled). Bottom: the nine planets with approximately correct relative sizes. (From NSSDC Photo Gallery.)

tion, and its consequences for the long-term evolution of the planetary system. The possible correlation between the chaotic changes of planetary orbits and the history of the climate of the terrestrial planets is still under investigation.

If the planets present very interesting dynamical behavior, even more interesting is the dynamics of the asteroids. Because of their large number, for each dynamical phenomenon, even the most pathological one, it is almost always possible to find an asteroid that exhibits it. The asteroids are small rocky or icy bodies; the biggest one, Ceres, is 900 km in diameter. Most of them are in the region between Mars and Jupiter, called the *main belt*, on orbits that do not intersect those of the planets.

If we represent the distribution of the asteroids using the semimajor axis, eccentricity and inclination, which characterize their elliptical orbits (see Chapter 1 for a precise definition of these *orbital elements*) some puzzling features immediately appear evident (Fig. 2). First of all, at some precise values of the semimajor axis (2.5, 2.8, 3.3 astronomical units (AU); the mean heliocentric distance of the Earth is 1 AU) there are well-defined gaps in the

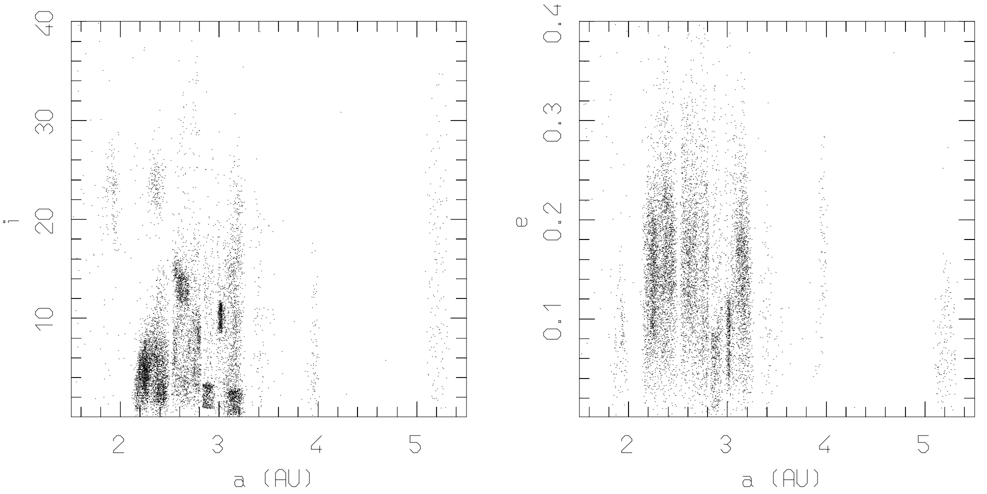


Figure 2: The orbital distribution of the first 10,000 asteroids in the space of osculating semimajor axis (a), eccentricity (e) and inclination (i).

asteroid distribution. These features are named *Kirkwood gaps*, from the name of the astronomer who first noticed them in 1866. The location of these gaps coincides with the location of the main *mean motion resonances* with Jupiter, which occur when the orbital period of the asteroid and of the planet are in integer ratio. However, understanding why asteroids dislike these Jovian resonances has been an open problem for more than a century. For some reason, the association between mean motion resonances and distribution gaps is not a general rule, because the concentration of the asteroids with semimajor axis close to 4 AU coincides with the location of the 3/2 resonance with Jupiter. Only recently has quite a complete solution for the problem of the origin of the Kirkwood gaps been provided.

However the Kirkwood gaps are not the only gaps in the asteroid distribution. For instance, the left plot of Fig. 2 shows another gap with a curved shape, approximately going from $a = 2$ AU, $i = 0°$ to $a = 2.5$ AU, $i = 18°$. This gap is associated with a resonance of a new type, called *secular resonance*. Secular resonance occurs when an integer combination of the precession rates of the asteroid's orbit is equal to an integer combination of the precession rates of the planetary orbits. Another secular resonance has been recently found to be responsible for the absence of asteroids between the two well-defined groups that are visible in Fig. 2 at $i \sim$20-28 degrees, $a \sim 1.9$ AU and $a \sim 2.3$ AU, respectively.

Figure 2 also shows some evident groupings of asteroids. It is tempting

to conjecture that the bodies in these groupings are genetically related (e.g. formed by the break-up of a unique parent body). However, a simple Celestial Mechanics computation shows that the semimajor axis, eccentricity and inclination oscillate with time, due to planetary perturbations. Thus, the visible groupings could in principle be just apparent features. On the other hand, there might exist genetically related groups of asteroids which are not visible in Fig. 2 because of these oscillations. A task of modern Celestial Mechanics is therefore to compute quantities that are quasi-invariant with time, called *proper elements*, which can be used to select all groups of asteroids (called *families*) that are statistically significant.

The asteroids with semimajor axis close to 5.2 AU are the so-called *Greeks* and *Trojans*. They have the same orbital period as Jupiter, and are approximately 60 degrees ahead and behind the planet, respectively. Lagrange showed that their orbital configuration is stable to a first approximation. However, accounting for the asteroids' large inclinations, Jupiter's eccentricity and the presence of the other planets, the stability of the Greeks and Trojans is far from trivial. A puzzling feature is that the other giant planets do not have Greek or Trojan asteroids, which might be a consequence of the quasi-resonances existing among these planets. Among the terrestrial planets, only Mars has two Trojan asteroids.

Several asteroids exist on orbits that undergo close encounters with the terrestrial planets. According to their current eccentricity and semimajor axis, they are named Apollos, Amors, Atens and Mars-crossers (Fig. 3). The first three classes constitute what is usually called the *near-Earth asteroid population* (NEAs).

Numerical integration shows that the typical lifetime of NEAs is 10 My. NEAs are eliminated by collision with the Sun, ejection from the Solar System, or collision with the planets. Understanding how this happens is another task of modern Celestial Mechanics that requires modeling the location and the strength of resonances as well as the effects of close encounters with the planets. The short lifetime of NEAs implies that this population is certainly not primordial, but it must be kept in a sort of steady state by a flux of asteroids that leave the main belt and start to cross the planetary orbits. It is intuitive that the origin of NEAs is related to the gaps in the main belt asteroid distribution. The collisions and the fragmentation of asteroids, as well as weak nonconservative forces, continuously emplace new bodies into the gaps; the rapid action of the main secular and mean motion resonances keeps the gaps clear, forcing the bodies to leave the main belt and to reach NEA-like orbits. However, the concentration of Mars-crossing asteroids between 2.1 and 2.5 AU suggests that many bodies escape also from the portion of the main

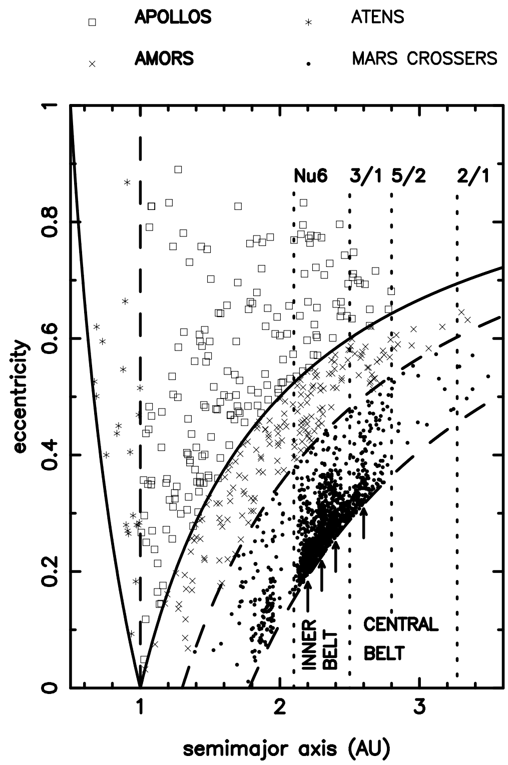


Figure 3: The orbital distribution of the known Apollos, Amors, Atens and Mars-crossing asteroids. The solid and dash curves mark the boundaries of these populations, while the short dash lines mark the location of some of the most important resonances. The arrows suggest that asteroids in the inner and the central part of the main asteroid belt might migrate in eccentricity, thus becoming Mars-crossers. The origin of near-Earth and Mars-crossing asteroids relies on the properties of resonant dynamics in the main belt.

belt that is located in the same semimajor axis range, although the latter is not characterized by any evident gap. It has been recently shown that this is due to the action of numerous weak resonances, whose destabilizing effects manifest themselves on timescales that are long enough to allow their complete replenishment.

Beyond the orbit of Neptune, there is a second belt of small bodies known as the *Kuiper belt.* As June 9, 2000, 279 trans-Neptunian objects have been discovered. The "planet" Pluto is also in the Kuiper belt. In 1999 it was intensely debated whether Pluto should still be considered as a planet, or should it be declassed to the status of a small object of the Solar System. The

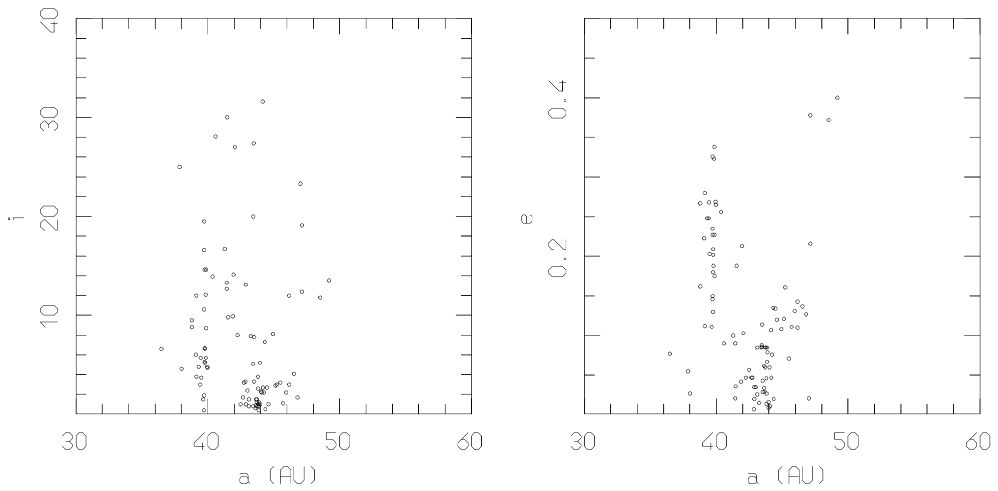


Figure 4: The orbital distribution of the Kuiper belt objects that have been observed during at least two oppositions. Data are taken from the Minor Planet Center list on June 9, 2000.

International Astronomical Union has decided to conserve Pluto's historical status, although many scientists agree that, for its orbit (much more eccentric and inclined than those of all other planets) and for its size (only 2340 km in diameter, 2/3 the size of the Moon), Pluto should be better regarded as the biggest object that has been discovered so far in the Kuiper belt.

Our knowledge of the orbital distribution of Kuiper belt bodies is still very limited because, since Pluto's discovery, the first object was found in 1992, and several years of observations are required in order to achieve a good determination of the orbital parameters. Figure 4 is the analog of Fig. 2, for the Kuiper belt objects that have been observed during at least two oppositions.

Despite the small total number of these objects (94), a few features in their orbital distribution recall the characteristics of the main asteroid belt. The concentration of bodies at 39.5 AU – among which Pluto is also located – is associated with the 2/3 mean motion resonance with Neptune. Moreover, a "gap" is visible in the 40-42 AU region, due to the presence of secular resonances. The outer bound of the Kuiper belt is not known. Strong biases act against the discovery of the farthest bodies, so that our knowledge is limited to the inner part of the Kuiper belt. One of the goals of Celestial Mechanics is to anticipate the discovery of Kuiper belt objects, and predict their orbital distribution on a theoretical basis.

If the asteroid belt is responsible for sustaining the NEA population, the

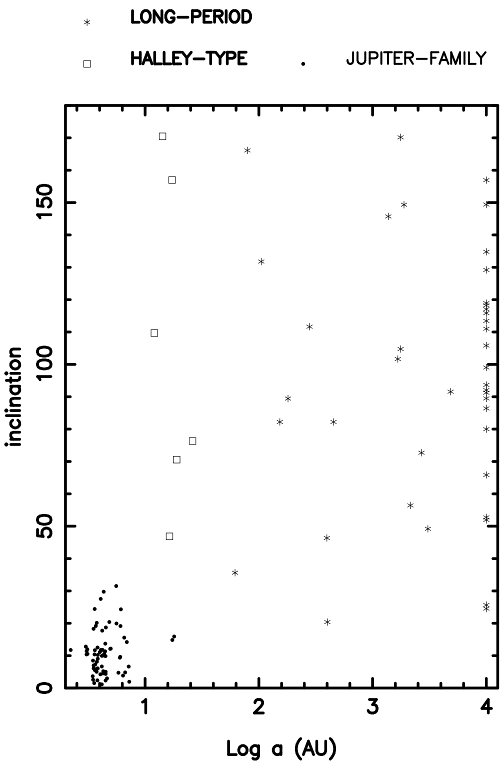


Figure 5: Orbital distribution and classification of comets. By convention, long-period comets are those with orbital period exceeding 200 years. The comets that have a shorter period are subdivided into the Jupiter family and Halley type groups, according to the relative speed by which they encounter Jupiter's orbit.

Kuiper belt is believed to be responsible for sustaining the population of the so-called *Jupiter-family comets.* These are low-inclined, short-period comets (Fig. 5), whose dynamics is dominated by encounters with Jupiter. In recent years, several studies have been devoted to the mechanisms by which Kuiper belt bodies start to encounter Neptune and are subsequently transported to the inner Solar System, where they are observed and categorized as Jupiter-family comets.

The long-period comets and the so-called *Halley-type* comets, by their isotropic distribution in inclination (Fig. 5), are not believed to come from a disk reservoir like the Kuiper belt, but rather form a quasi-spherical reservoir, called the *Oort cloud.* The surplus of comets with semimajor axis of order 10,000 AU suggests that this reservoir is at the frontier of the Solar Sys-

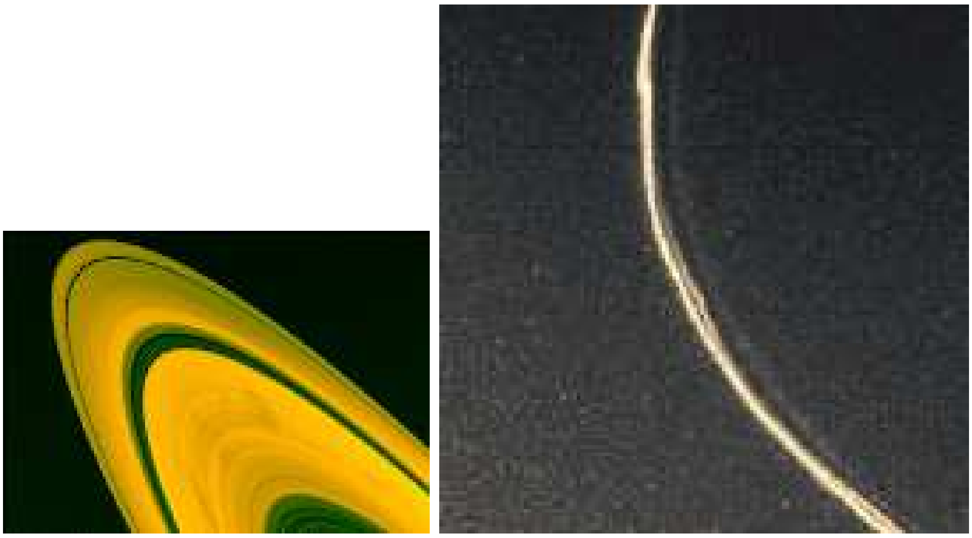


Figure 6: Left: global view of Saturn's rings, showing the Cassini and Encke divisions, respectively closer and further from the planet. Right: a detail of the complex structure of the F-ring of Saturn. (From NSSDC Photo Gallery.)

tem. It is generally believed that the Oort cloud was formed by planetesimals originally in the giant-planet region, that were scattered by close encounters with the planets onto very elliptic orbits, with semimajor axis ~10,000 AU. At such a distance from the Sun, the gravitational potential of the entire galaxy becomes a strong perturbation of the central potential of the Sun. The study of the dynamical effect of this *Galactic tide* is crucial to understand the formation of the Oort cloud, and the sporadic re-injection of comets into the inner Solar System. Models taking into account the Galactic tide, the passage of stars and of giant molecular clouds, aim at determining how many comets should be stocked in the Oort cloud in order to account for the observed number of long-period comets. This result in turn allows the estimate of the total mass of the original planetesimal disk.

The ring systems of Jupiter, Saturn, Uranus and Neptune provide wonderful natural laboratories in which to study the effects of both interparticle and external gravitational forces, combined with the effects of physical collisions. The magnificent system of Saturn's rings (see the left picture in Fig. 6) shows two wide gaps: the 4500 km wide Cassini division, between the A and B rings, and the 325 km wide Encke division, within the A ring. These gaps in the distribution of the ring particles recall the Kirkwood gaps in the asteroid belt. In fact, they also coincide with the location of mean motion resonances

with some of the satellites of Saturn, or to the overlapping of a system of resonances. For instance, the Cassini division is determined by the 2/1 resonance with Mimas, while the outer edge of the A ring is produced by the 7/6 resonance with the co-orbital satellites Janus and Epimetheus.

In addition to the Cassini and the Encke divisions, several other narrower gaps are visible in close-up images of Saturn's rings. Most of these divisions have been understood to be associated with resonances with satellites, sometimes involving the ratios of large integer numbers, as in the case of the 32/31 resonance with Prometheus, at the location of the so-called *Keeler gap*.

Some of the gaps are not completely empty, but host narrow *ringlets*, typically a few km wide. Unlike the main rings, all these ringlets are decidedly noncircular, raising the problem of the dynamical mechanisms responsible for their atypical shape. Saturn's F-ring (see the right picture in Fig. 6), is also a complex structure made up of two narrow, braided, bright rings along which "knots" are visible. Scientists speculate that the knots may be clumps of ring material, or mini-moons. The shape of this ring seems to be dictated by the presence of the so-called *shepherding satellites*, orbiting along each of its sides.

Finally, the systems of satellites of the giant planets can be considered as miniature solar systems, also rich in intriguing dynamical phenomena. Mean motion resonances are very common, because the tides exerted by the planet force the satellites' slow outward migration. Because the relative orbital periods change with the semimajor axes, the satellites must pass through several resonant configurations. In some cases, a resonance traps the satellites, so that their semimajor axes expand in such a way that the ratio between their orbital periods remains unchanged. Particularly interesting is the so-called *Laplace resonance*, which is a triple mean motion resonance, among the Galilean satellites Io, Europa and Ganymede: Io orbits twice around Jupiter for each orbit of Europa, which in turn orbits twice for each orbit of Ganymede.

The interplay between tidal forces and orbital dynamics also explains why the eccentricities of many satellites close to the planets are not equal to zero, despite the rapid tidal damping. The same is true also for the satellites' inclinations: for instance, the large inclination of the Uranian satellite Miranda (4°) is explained by a temporary trapping in the 1/3 resonance with Umbriel, which occurred in the recent past.

The tides also tend to slow down the spin frequency of the satellites, until the rotational period becomes synchronous with the orbital period. In this case, the satellite always faces the same side toward its planet, like Earth's Moon. However, in some cases the satellite may enter chaotic zones, becoming attitude-unstable and tumbling. This is for instance the case of Saturn's satellite Hyperion, one of the first examples of chaotic motion found in our

Figure 7: The Jovian satellite Io. It is the only Solar System body other than the Earth showing volcanic activity. The internal heating of the satellite is sustained by tidal interactions between Io, Europa, Ganymede and Jupiter. (From NSSDC Photo Gallery.)

Solar System.

The internal heating of several satellites is due to the interplay between orbital eccentricity and synchronous rotation. The most magnificent example is that of the Jovian satellite Io, which is the only body of the Solar System, other than the Earth, exhibiting active volcanism. Several other satellites show evidence of recent tectonic activity, and Europa could still have a liquid water ocean under its surface.

Chapter 1

ELEMENTARY CELESTIAL AND HAMILTONIAN MECHANICS

1.1 Equations of motion

According to Newton's theory of gravitation, the equations of motion of an isolated system of two bodies having spherical symmetry and masses m_0 and m_1 are:

$$\frac{\mathrm{d}^2\mathbf{u}_0}{\mathrm{d}t^2} = \frac{\mathcal{G}m_1}{\|\mathbf{u}_1 - \mathbf{u}_0\|^3}(\mathbf{u}_1 - \mathbf{u}_0) \ , \quad \frac{\mathrm{d}^2\mathbf{u}_1}{\mathrm{d}t^2} = \frac{\mathcal{G}m_0}{\|\mathbf{u}_0 - \mathbf{u}_1\|^3}(\mathbf{u}_0 - \mathbf{u}_1) \ , \tag{1.1}$$

where $\mathbf{u}_0$ and $\mathbf{u}_1$ denote the position vectors of the two bodies in an inertial reference frame (also called *barycentric coordinates* if the origin is in the barycenter of the system), $\mathcal{G}$ is the gravitational constant and $\|\mathbf{x}\|$ denotes the Euclidean norm of a generic vector $\mathbf{x}$.

Denoting by $\mathbf{r} = \mathbf{u}_1 - \mathbf{u}_0$ the relative position of the bodies (r_1, r_2, r_3 are called *heliocentric coordinates* if the body at $\mathbf{u}_0$ is the Sun), the two vectorial equations above can be reduced to two separate vectorial equations:

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\frac{\mathcal{G}(m_0 + m_1)}{\|\mathbf{r}\|^3}\mathbf{r} \ , \tag{1.2}$$

that describes the relative motion of the two bodies and

$$\frac{\mathrm{d}^2\mathbf{s}}{\mathrm{d}t^2} = 0 \ , \tag{1.3}$$

which shows that the barycenter of the system, with coordinate vector $\mathbf{s} = m_0\mathbf{u}_0 + m_1\mathbf{u}_1$, keeps its inertial motion.

Equation (1.2) can be solved exactly, the solution depending only on the initial relative position $\mathbf{r}(0)$ and velocity $\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}(0)$. A detailed explanation of how the solution is achieved can be found in Chapter 6 of Danby (1962). The solution of equation (1.2) was the first great success of Newton's theory of gravitation, because of its perfect agreement – for bounded motion – with the three laws that Kepler empirically deduced from observations of the planet Mars:

Law 1: Each planet moves, relative to the Sun, in an elliptical orbit, the Sun being at one of the two foci of the ellipse.

Law 2: The rate of motion in the elliptical orbit is such that the vector pointing to the position of the planet relative to the Sun spans equal areas of the orbital plane in equal times.

Law 3: The square of the orbital period T is proportional to the cube of the semimajor axis a of the orbital ellipse.

Therefore, it is convenient for astronomers to characterize the relative motion of the two bodies by quantities that describe the geometrical properties of the orbital ellipse and the instantaneous position on the ellipse. These quantities are usually called *orbital elements*. Remember that the ellipse describes the motion of one body relative to another one, usually denoted respectively as the *secondary* and the *central* bodies. The choice of the central body is arbitrary: the same ellipse describes the motion of both m_0 around m_1 and of m_1 around m_0. Going back to an inertial reference frame using the solution of (1.3), one easily sees that the bodies describe similar elliptical trajectories around their common baricenter; the ratio between the sizes of the two ellipses is inversely proportional to the mass ratio. As a consequence, when the mass ratio tends to zero, as in the case of a small body and a star, the star's orbit shrinks to the baricenter's position, and the orbit relative to the star becomes equal to the small body's orbit in the inertial reference frame. For this reason, although from the mathematical viewpoint the choice is arbitrary, it is reasonable to choose as the central body the more massive of the two bodies.

1.2 Orbital elements

Let us first define quantities that describe the shape of the orbital ellipse around the central body and the position of the secondary body on the ellipse.

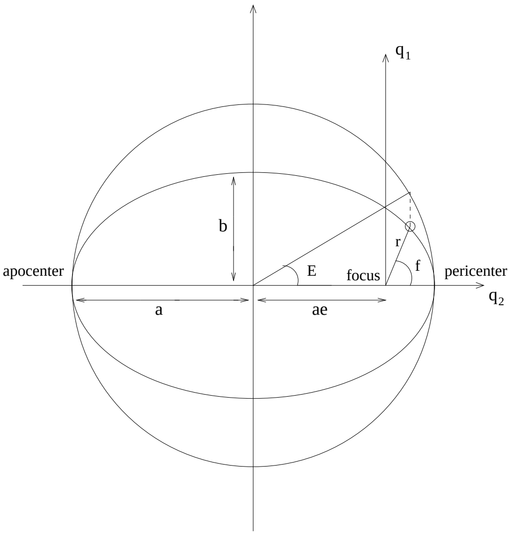


Figure 1.1: Keplerian motion: definition of a, e and E.

Then we will introduce quantities describing the orientation of the ellipse in space.

The shape of the ellipse can be completely determined by two quantities: the semimajor axis a and semiminor axis b, or, more suitably, the semimajor axis a and the eccentricity e (Fig. 1.1). The name *eccentricity* comes from e being the ratio between the distance of the focus from the center of the ellipse and the semimajor axis of the ellipse. The eccentricity is therefore an indicator of how much the orbit differs from a circular one: $e = 0$ means that the orbit is circular, while $e = 1$ means that the orbit is a segment of length $2a$, the central body being at one of the extremes. Among all "elliptical" trajectories, the latter is the only collisional one, if the physical radii of the bodies are neglected. A semimajor axis of $a = \infty$ and $e = 1$ denote parabolic motion, while the convention $a < 0$ and $e > 1$ is adopted for hyperbolic motion. We will rarely deal with these kinds of unbounded motion in this book, hence we will concentrate, hereafter, on the elliptical case. On an elliptical orbit, the closest point to the central body is called the *pericenter* (or alternatively *perihelion* if the central body is the Sun, or *perigee* if the central body is the Earth), and its distance q is equal to $a(1 - e)$; the farthest point is called the *apocenter* (respectively *aphelion* or *apogee*), and its distance Q is equal to $a(1 + e)$.

To denote the position of the body on the elliptic orbit it is convenient to use an orthogonal reference frame q_1, q_2 with origin at the focus of the ellipse

occupied by the central body and q_1 axis oriented towards the pericenter of the orbit. Alternatively, polar coordinates r, f can be used. The angle f is usually called the *true anomaly* of the body. From Fig. 1.1, with elementary geometrical relationships one has

$$q_1 = a(\cos E - e)\ , \quad q_2 = a\sqrt{1-e^2}\sin E \tag{1.4}$$

and

$$r = a(1 - e\cos E)\ , \quad \cos f = \frac{\cos E - e}{1 - e\cos E} \tag{1.5}$$

where E is the angle subtended at the center of the ellipse by the projection of the position of the body on the circle with radius a and tangent to the ellipse at pericenter and apocenter, as Fig. 1.1 shows. This angle is called the *eccentric anomaly.* The position of a body in its orbit can be expressed in terms of a, e and E only.

From the equations of motion, it is possible to derive (Danby, 1962) the evolution law of E with respect to time, usually called the *Kepler equation*:

$$E - e\sin E = n(t - t_0) \tag{1.6}$$

where

$$n = \sqrt{\mathcal{G}(m_0 + m_1)}a^{-3/2} \tag{1.7}$$

is the orbital frequency or *mean motion* of the body (in agreement with Kepler's second law), t is the time and t_0 is the time of passage at pericenter.

Astronomers like to introduce a new angle

$$M = n(t - t_0) \tag{1.8}$$

called the *mean anomaly,* as an orbital element that changes linearly with time and still denotes the position of the body in its orbit (through equations (1.6) and (1.5)).

To characterize the orientation of the ellipse in space, with respect to an arbitrary orthogonal reference frame (x, y, z) centered on the position of the central body, we have to introduce three additional angles (see Fig. 1.2). The first one is the inclination i of the orbital plane (the plane which contains the ellipse) with respect to the (x, y) reference plane. If the orbit has a nonzero inclination, it intersects the (x, y) plane in two points, called the *nodes* of the orbit. Astronomers distinguish between an *ascending* node, where the body passes from negative to positive z and a *descending* node, where the body plunges towards negative z. The orientation of the orbital plane in space is then completely determined when the angular position of the ascending node from the x axis is given. This angle is traditionally called the *longitude of node*,

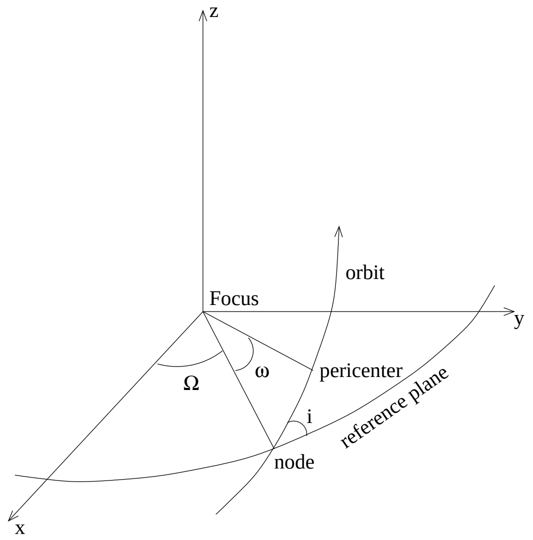


Figure 1.2: Keplerian motion: definition of i, Ω and ω.

and is usually denoted by Ω. The last angle that needs to be introduced is the one characterizing the orientation of the ellipse in its plane. The *argument of pericenter* ω (also called the *argument of perihelion* if the central body is the Sun) is defined as the angular position of the pericenter, measured in the orbital plane relative to the line connecting the central body to the ascending node.

The orbital elements a, e, i, ω, Ω and M completely define the position and velocity of the secondary body with respect to the central one. There is a one-to-one correspondence between $x, y, z, \mathrm{d}x/\mathrm{d}t, \mathrm{d}y/\mathrm{d}t, \mathrm{d}z/\mathrm{d}t$ – components of $\mathbf{r}$ and $\mathrm{d}\mathbf{r}/\mathrm{d}t$ – and the orbital elements. This correspondence is given by the relationship:

$$\mathbf{r} = R_{xq}\mathbf{q} \ , \quad \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = R_{xq}\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} \ , \tag{1.9}$$

where the vectors $\mathbf{q}$ and $\mathrm{d}\mathbf{q}/\mathrm{d}t$ have components

$$[q_1, q_2, 0] = [a(\cos E - e), a\sqrt{1-e^2}\sin E, 0] \tag{1.10}$$

and

$$\left[\frac{\mathrm{d}q_1}{\mathrm{d}t}, \frac{\mathrm{d}q_2}{\mathrm{d}t}, 0\right] = \left[-\frac{na\sin E}{1-e\cos E}, \frac{na\sqrt{1-e^2}\cos E}{1-e\cos E}, 0\right] \tag{1.11}$$

respectively, and the rotation matrix R_{xq} has entries

$$\begin{pmatrix} \cos\Omega\cos\omega - \sin\Omega\cos i\sin\omega & -\cos\Omega\sin\omega - \sin\Omega\cos i\cos\omega & \sin\Omega\sin i \\ \sin\Omega\cos\omega + \cos\Omega\cos i\sin\omega & -\sin\Omega\sin\omega + \cos\Omega\cos i\cos\omega & -\cos\Omega\sin i \\ \sin i\sin\omega & \sin i\cos\omega & \cos i \end{pmatrix} . \tag{1.12}$$

In the definition of the orbital elements above, note that when the inclination is zero, ω and M are not defined, because the position of the ascending node is not determined. Moreover, M is not defined also when the eccentricity is zero, because the position of the pericenter is not determined. It is convenient, therefore, to introduce the *longitude of perihelion* $\varpi = \omega + \Omega$ and the *mean longitude* $\lambda = M + \omega + \Omega$. The first angle is well defined when $i = 0$, while the second one is well defined when $i = 0$ and/or $e = 0$. It is evident that also the set of orbital elements $a, e, i, \varpi, \Omega, \lambda$ unequivocally defines the position and velocity of the body. Note that, setting $i = 0$, the relations (1.12) become in a natural way dependent on ϖ only (and not separately on Ω and ω); similarly, setting $e = 0$, (1.9) become dependent on $\lambda - \Omega$ (and not separately on ω and E).

Finally, from (1.2) and using (1.9) it is possible to check that equation (1.2) preserves "energy"

$$\mathcal{H} = \frac{1}{2}\left\|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right\|^2 - \frac{\mathcal{G}(m_0 + m_1)}{\|\mathbf{r}\|} = -\frac{\mathcal{G}(m_0 + m_1)}{2a} \tag{1.13}$$

and "angular momentum" $\mathbf{r} \times \mathrm{d}\mathbf{r}/\mathrm{d}t$, whose norm and projection on the z axis are respectively:

$$G = \sqrt{\mathcal{G}(m_0 + m_1)a(1 - e^2)} \ , \tag{1.14}$$

$$H = G\cos i \ . \tag{1.15}$$

These relations will be important in Section 1.9.1.

1.3 Perturbations of the two-body problem

In a barycentric inertial reference frame, the equations of motion of an isolated system made of a Sun of mass m_0 and of N planets of masses $m_1, m_2, \ldots, m_N$ are:

$$\frac{\mathrm{d}^2\mathbf{u}_i}{\mathrm{d}t^2} = -\mathcal{G}\sum_{j\neq i} m_j \frac{\mathbf{u}_i - \mathbf{u}_j}{\|\mathbf{u}_i - \mathbf{u}_j\|^3} \ , \tag{1.16}$$

where i and j range from 0 to N. Introducing the heliocentric positions of the planets $\mathbf{r}_i = \mathbf{u}_i - \mathbf{u}_0$ the equations above can be rewritten as

$$\frac{\mathrm{d}^2\mathbf{r}_i}{\mathrm{d}t^2} = -\frac{\mathcal{G}(m_0+m_i)}{\|\mathbf{r}_i\|^3}\mathbf{r}_i + \sum_{j=1, j\neq i}^{N} \mathcal{G}m_j \left(\frac{\mathbf{r}_j - \mathbf{r}_i}{\|\mathbf{r}_j - \mathbf{r}_i\|^3} - \frac{\mathbf{r}_j}{\|\mathbf{r}_j\|^3} \right) , \quad i = 1, \ldots, N , \tag{1.17}$$

while the motion of the Sun is given by $\mathbf{u}_0 = -\sum_{i=1}^N m_i \mathbf{r}_i / \sum_{i=0}^N m_i$. Equations (1.17) constitute what is usually called the $(N+1)$-*body problem.*

Similarly, the heliocentric equation of motion of a body of negligible mass under the gravitational forces of the Sun and of N planets in *given* orbits is

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\frac{\mathcal{G}m_0}{\|\mathbf{r}\|^3}\mathbf{r} + \sum_{j=1}^{N} \mathcal{G}m_j \left(\frac{\mathbf{r}_j - \mathbf{r}}{\|\mathbf{r}_j - \mathbf{r}\|^3} - \frac{\mathbf{r}_j}{\|\mathbf{r}_j\|^3} \right) , \tag{1.18}$$

where $\mathbf{r}$ is the position vector of the small body with respect to the Sun and $\mathbf{r}_j$ are the heliocentric positions of the planets of masses m_j. Equation (1.18) constitutes what is usually called the *restricted problem* (*restricted three-body problem* if only one planet is considered).

If the masses of the planets are small compared to that of the Sun, and none of their mutual distances $\mathbf{r}_j - \mathbf{r}_i$ becomes small, equations (1.17) and (1.18) are evidently close to the equation of the two-body problem (1.2), the term dependent on the planetary masses m_j playing the role of a small *perturbation* with respect to the two-body interaction with the Sun. Therefore, the motion resulting from equations (1.17) and (1.18) will be close to Keplerian motion. As a consequence, rewriting equations (1.17) and (1.18) in terms of the orbital elements using the relations (1.9), the equations for a, e, i, ω, Ω have the form $\mathrm{d}\alpha/\mathrm{d}t = O(m_j/m_0)$, where α denotes any of these elements and $O(m_j/m_0)$ denotes a function which is as small as the mass of the planets in solar mass units; the equation for M has the form $\mathrm{d}M/\mathrm{d}t = n + O(m_j/m_0)$, where n is the unperturbed mean motion resulting from the two-body problem. These equations of motion for the orbital elements are usually called the *Lagrange equations*, and show that the orbital elements a, e, i, ω, Ω change slowly with time, while M deviates slowly from its linear unperturbed motion. For this reason, it is still convenient to characterize the planets and the small bodies of the Solar System by their orbital elements knowing that, in general, these elements will change very little on the timescale of human life (this is indeed the reason why Kepler discovered the laws of the two-body problem, although he was observing the motions of real planets). However, it should not be forgotten that the orbital elements do change with time. Strictly speaking, they only represent the instantaneous orbit that the body would have if all perturbations suddenly disappeared. In more mathematical terms, the orbital

elements at time t represent the Keplerian motion that is tangent to the real motion that the body has at time t. For this reason they are also known as *osculating elements*.[1]

The Lagrange equations are difficult to handle, and we will have to rewrite them in a slightly different form – a Hamiltonian form – before being able to study the motion in detail.

1.4 Hamiltonian systems and the two-body problem

A system of ordinary differential equations of the type

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{F}(\mathbf{r}) \tag{1.19}$$

is said to be in *Hamiltonian form* if $\mathbf{r}$ is a $2n$-uple and, denoting by $x_1, \ldots, x_n$ and $v_1, \ldots, v_n$ its $2n$ components, there exists a function $\mathcal{H}(v_1, \ldots, v_n, x_1, \ldots, x_n)$ such that equations (1.19) can be rewritten as

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = -\frac{\partial \mathcal{H}}{\partial x_i} \, , \quad \frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial v_i} \, , \tag{1.20}$$

for $i = 1, \ldots, n$. The function $\mathcal{H}$ is called the *Hamiltonian* of the system and the variables $x_1, \ldots, x_n$ and $v_1, \ldots, v_n$ are respectively called *coordinates* and *momenta*. For simplicity, we will adopt the vectorial formalism, denoting by $\mathbf{x}$ and $\mathbf{v}$ the n-uples $x_1, \ldots, x_n$ and $v_1, \ldots, v_n$. Moreover, we will denote hereafter by $\dot{\alpha}$ the time derivative of a generic variable α. $\mathbf{v}$ and $\mathbf{x}$ are called *conjugate variables* and the $(\mathbf{x}, \mathbf{v})$ space is usually called the *phase space* of the system. The dimension n of the vectors $\mathbf{x}$ and $\mathbf{v}$ is called the *number of degrees of freedom.*

A large class of ordinary differential equations, and in particular most of the equations that concern Celestial Mechanics, can be written in Hamiltonian form. In fact, equations of second order of type

$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = -\mathrm{grad}_{\mathbf{r}} U(\mathbf{r}) \tag{1.21}$$

can be rewritten in the form (1.20), as is straightforward to check, by setting $x_i = r_i$, $v_i = \dot{r}_i$ and $\mathcal{H} = 1/2\|\mathbf{v}\|^2 + U(\mathbf{x})$. In particular, the relative equations

[1]This property is not true for the *formal osculating elements* introduced in Section 1.9.2, which define an ellipse that transversally intersects the real trajectory.

of the two-body problem (1.2) are in Hamiltonian form with

$$\mathcal{H} = \frac{v_1^2 + v_2^2 + v_3^2}{2} - \frac{\mathcal{G}(m_0 + m_1)}{\sqrt{x_1^2 + x_2^2 + x_3^2}} \tag{1.22}$$

where $x_1, \ldots, x_3$ are the usual Cartesian coordinates of a reference frame centered on the central body and $v_i = \dot{x}_i$ are the speeds.

More generally, the systems that conserve a quantity – say the "energy" – which in the variables $\mathbf{r}$ and $\dot{\mathbf{r}}$ can be written as the sum of a "kinetic energy" $T(\mathbf{r}, \dot{\mathbf{r}})$ and of a "potential energy" $U(\mathbf{r})$, admit a Hamiltonian form.[2] Decomposing $T = T_2 + T_1 + T_0$, where T_2, T_1 and T_0 are respectively the terms of order 2, 1 and 0 in $\dot{\mathbf{r}}$, the equations of motion can be written as in (1.20), with

$$x_i = r_i \ , \quad v_i = \frac{\partial T}{\partial \dot{r}_i} \ , \quad \mathcal{H} = T_2 - T_0 + U \ . \tag{1.23}$$

(see Whittaker (1937), Section 7, for a derivation).

The above recipe is very useful for writing the Hamiltonian of a system in coordinates such that the equations are not straightforwardly in the form (1.21), as is the case, for instance, when spherical coordinates are used, or a rotating reference frame is chosen.

As a first example, consider the two-body problem (1.2), but in the usual spherical coordinates $\varrho, \vartheta, \varphi$. Recalling the relationships with the Cartesian coordinates $\mathbf{r} \equiv (x, y, z)$, which are $x = \varrho \sin\vartheta \cos\varphi$, $y = \varrho \sin\vartheta \sin\varphi$, $z = \varrho \cos\vartheta$, the "potential energy" can be written $U(\varrho) = -\mathcal{G}(m_0 + m_1)/\varrho$, while the "kinetic energy" is $T = 1/2[\dot{\varrho}^2 + \varrho^2\dot{\vartheta}^2 + \varrho^2 \sin^2\vartheta\dot{\varphi}^2]$. The Hamiltonian of the two-body problem in spherical coordinates will then be:

$$\mathcal{H} = \frac{1}{2}\left(p_\varrho^2 + \frac{p_\vartheta^2}{\varrho^2} + \frac{p_\varphi^2}{\varrho^2 \sin^2\vartheta}\right) - \frac{\mathcal{G}(m_0 + m_1)}{\varrho} \tag{1.24}$$

with coordinates r, ϑ and φ and momenta $p_\varrho = \dot{\varrho}$, $p_\vartheta = \varrho^2\dot{\vartheta}$ and $p_\varphi = \varrho^2 \sin^2\vartheta\dot{\varphi}$.

As a second example, consider the two-body problem in Cartesian coordinates, rotating with angular speed ω around the z axis. Denoting by X, Y, Z the Cartesian coordinates in an inertial reference frame, and by x, y, z the coordinates in the rotating reference frame, using the relationships $X = x\cos\omega t - y\sin\omega t$, $Y = x\sin\omega t + y\cos\omega t$, and $Z = z$, it is straightforward to see that the "kinetic energy" $T = 1/2[\dot{X}^2 + \dot{Y}^2 + \dot{Z}^2]$ becomes

[2] It is not necessary that the conserved quantity is equal to the real energy of the system. In the first two examples below, for instance, we consider an "energy" that is the real energy divided by the mass m_1.

$T = 1/2[\dot{x}^2 + \dot{y}^2 + \dot{z}^2 + \omega^2(x^2 + y^2) - 2\omega(\dot{x}y - x\dot{y})]$. Then, using the recipe above, introducing the momenta $p_x = \dot{x} - \omega y$, $p_y = \dot{y} + \omega x$, $p_z = \dot{z}$, the Hamiltonian of the two-body problem in the rotating frame can be written as

$$\mathcal{H} = \frac{1}{2}(p_x^2 + p_y^2 + p_z^2) + \omega(p_x y - x p_y) - \frac{\mathcal{G}(m_0 + m_1)}{\sqrt{x^2 + y^2 + z^2}} \ . \tag{1.25}$$

Finally, still using the same recipe, it is possible to see that the two-body problem, in the case $m_1 \neq 0$, admits also the Hamiltonian

$$\mathcal{H} = \frac{\|\tilde{\mathbf{v}}\|^2}{2\mu_1} - \frac{\mathcal{G}(m_0 + m_1)\mu_1}{\|\mathbf{r}\|} \tag{1.26}$$

in Cartesian coordinates $\mathbf{r}$ and conjugate momenta $\tilde{\mathbf{v}} = \mu_1 \dot{\mathbf{r}}$, where $\mu_1 = m_0 m_1/(m_0 + m_1)$ is the reduced mass of the secondary body. In fact, if one multiplies both sides of equations (1.2) by μ_1, the kinetic and potential energies of the system are

$$T = \frac{1}{2}\mu_1 \|\dot{\mathbf{r}}\|^2 \ , \quad U = -\frac{\mathcal{G}(m_0 + m_1)\mu_1}{\|r\|} \ , \tag{1.27}$$

from which, defining the momenta $\tilde{\mathbf{v}} = \partial T/\partial \dot{\mathbf{r}}$ and $\mathcal{H} = T + U$, (1.26) is obtained.

Of course, the Hamiltonians (1.22) (1.24) (1.25) and (1.26) are completely equivalent, in the sense that they all lead to the same equations of motion, although in different variables. These examples stress that the Hamiltonian of a given system is not unique, but depends on the choice of coordinates and momenta. Choosing one Hamiltonian rather than the others is just a matter of convenience.

1.5 Perturbations in Hamiltonian form

The equations (1.18) of the restricted problem can also be easily written in Hamiltonian form. It is enough to realize that the right-hand side of (1.18) is of the form $-\mathrm{grad}_{\mathbf{r}} U(\mathbf{r})$ with

$$U(\mathbf{r}) = -\frac{\mathcal{G} m_0}{\|\mathbf{r}\|} - \mathcal{G} \sum_{j=1}^{N} m_j \left(\frac{1}{\|\mathbf{\Delta}_j\|} - \frac{\mathbf{r} \cdot \mathbf{s}_j}{\|\mathbf{s}_j\|^3} \right) \ , \tag{1.28}$$

where $\mathbf{s}_j$ is the heliocentric position vector of the j-th planet and $\mathbf{\Delta}_j = \mathbf{r} - \mathbf{s}_j$; moreover in the above formula $\mathbf{r} \cdot \mathbf{s}_j$ denotes the scalar product between $\mathbf{r}$ and

$\mathbf{s}_j$. The Hamiltonian of the restricted problem is therefore

$$\mathcal{H} = \frac{\|\mathbf{v}\|^2}{2} - \frac{\mathcal{G}m_0}{\|\mathbf{r}\|} - \mathcal{G}\sum_{j=1}^{N} m_j \left(\frac{1}{\|\mathbf{\Delta}_j\|} - \frac{\mathbf{r}\cdot\mathbf{s}_j}{\|\mathbf{s}_j\|^3} \right) , \tag{1.29}$$

with Cartesian coordinates r_1, r_2, r_3 and momenta v_1, v_2, v_3 such that $\mathbf{v} = \dot{\mathbf{r}}$.

Note that the Hamiltonian (1.29) is of the form $\mathcal{H}_0 + \mathcal{H}_1$, where $\mathcal{H}_0 = \|\mathbf{v}\|^2/2 - \mathcal{G}m_0/\|\mathbf{r}\|$ is identical to the Hamiltonian (1.22) of the two-body problem with $m_1 = 0$, and $\mathcal{H}_1$ plays the role of a *perturbation*, whose size relative to $\mathcal{H}_0$ is proportional to the mass of the planets relative to that of the Sun. Moreover, in (1.29) the planets are assumed to move in given orbits, so that the vectors $\mathbf{s}_j$ are given functions of time. The Hamiltonian (1.29) has therefore the form $\mathcal{H}(\mathbf{v}, \mathbf{r}, t)$, and is an example of a *time-dependent Hamiltonian* or *nonautonomous Hamiltonian*, in contrast with the Hamiltonian of the two body problem that does not depend on time.

Nonautonomous Hamiltonians can always be transformed into autonomous Hamiltonians by extending the phase space. More precisely, one introduces a new coordinate τ, a conjugate momentum $\mathcal{T}$, and a new Hamiltonian

$$\mathcal{H}' = \mathcal{T} + \mathcal{H}(\mathbf{v}, \mathbf{r}, \tau) . \tag{1.30}$$

It is trivial to check that the equation for τ is $\dot{\tau} = \partial\mathcal{H}'/\partial\mathcal{T} = 1$, which gives the solution $\tau(t) = t$, so that the equations of motion for $\mathbf{v}$ and $\mathbf{r}$ given by the Hamiltonian $\mathcal{H}'$ are identical to those given by the Hamiltonian $\mathcal{H}$. With this trick the new Hamiltonian turns out to be formally independent of time, so that we will be able to apply the known results on autonomous Hamiltonian systems. As a matter of fact, when the time dependence has a complicated form, with several independent frequencies involved as in the case where several planets are considered, each with its own orbital frequency, it is convenient to overextend the phase space by introducing a coordinate τ_j and a conjugate momentum $\mathcal{T}_j$ for each independent frequency ν_j of the time-dependent perturbation. The Hamiltonian $\mathcal{H}(\mathbf{v}, \mathbf{r}, t)$ can then be rewritten as

$$\mathcal{H}' = \sum_j \nu_j \mathcal{T}_j + \mathcal{H}(\mathbf{v}, \mathbf{r}, \tau_1, \ldots, \tau_j) . \tag{1.31}$$

Again, it is trivial to check that the equations of motion for $\mathbf{v}$ and $\mathbf{r}$ are unchanged; the advantage of expression (1.31) with respect to (1.30) is that the coordinates τ_j are "angles", in the sense that the Hamiltonian is periodic in each of the τ_j. For instance, in the case of the restricted problem, the angles τ_j will be identified with the mean anomalies of the planets. This ensures the possibility of expanding the Hamiltonian in a Fourier series of τ_j and applying the machinery of perturbation techniques, as will be explained in Chapter 2.

We now come to the Hamiltonian of a planetary system. Equations (1.17) cannot be written in the form (1.21), so that we have to refer to equations (1.16) in an inertial reference frame. After multiplying both sides of (1.16) by m_i, it is easy to see that the right-hand side can be written as $-\text{grad}_{\mathbf{r}_i} U(\mathbf{r}_0, \ldots, \mathbf{r}_N)$, so that the "energy" of the system is:

$$\mathcal{H} = T + U = \frac{1}{2}\sum_{j=0}^{N} m_j \|\dot{\mathbf{u}}_j\|^2 - \mathcal{G}\sum_{j=1}^{N}\sum_{i=0}^{j-1} \frac{m_i m_j}{\|\mathbf{u}_i - \mathbf{u}_j\|} \ . \tag{1.32}$$

Therefore, as shown in the previous section, $\mathcal{H}$ is also the Hamiltonian of the system, once the momenta $\tilde{\mathbf{u}}_i = \partial T/\partial \dot{\mathbf{u}}_i = m_i \dot{\mathbf{u}}_i$ are introduced. It would be tempting to transform (1.32) by introducing $\mathbf{r}_i = \mathbf{u}_i - \mathbf{u}_0$ and $\mathbf{v}_i = \tilde{\mathbf{u}}_i - \tilde{\mathbf{u}}_0$ in order to use a heliocentric coordinate system. However, not all possible changes of coordinates and momenta (including this one) preserve the Hamiltonian form of the equations of motion. This leads us to introduce the concept of *canonical transformations.*

1.6 Canonical transformations

A time-independent transformation of coordinates and momenta in phase space is said to be canonical if it preserves the Hamiltonian form of the equations of motion, whatever the Hamiltonian function. More precisely, a transformation $(\mathbf{v}, \mathbf{x}) \to (\mathbf{v}', \mathbf{x}')$ is canonical if, $\forall \mathcal{H}$, defining

$$\mathcal{H}'(\mathbf{v}', \mathbf{x}') = \mathcal{H}(\mathbf{v}(\mathbf{v}', \mathbf{x}'), \mathbf{x}(\mathbf{v}', \mathbf{x}')) \tag{1.33}$$

the equations of motion for $\mathbf{v}'$ and $\mathbf{x}'$ become:

$$\dot{\mathbf{v}}' = -\frac{\partial \mathcal{H}'}{\partial \mathbf{x}'} \ , \quad \dot{\mathbf{x}}' = \frac{\partial \mathcal{H}'}{\partial \mathbf{v}'} \ . \tag{1.34}$$

The class of canonical transformation is a very restrictive class among all possible transformations on the phase space. Nevertheless, we are compelled to restrict to canonical transformation if we don't want to lose the Hamiltonian form of the equations, which will be so useful to allow a detailed study of the motion. It is therefore important to have criteria to test whether a given transformation is canonical. Without trying to give an exhaustive list, we discuss in the following three criteria that will be used later in this book.

The most useful criterion is that of *Poisson brackets.* Given conjugate coordinates and momenta $\mathbf{x}$ and $\mathbf{v}$ and functions $f(\mathbf{v}, \mathbf{x})$ and $g(\mathbf{v}, \mathbf{x})$, the Poisson bracket between f and g is defined as

$$\{f, g\} = \text{grad}_{\mathbf{x}} f \cdot \text{grad}_{\mathbf{v}} g - \text{grad}_{\mathbf{v}} f \cdot \text{grad}_{\mathbf{x}} g = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial g}{\partial v_i} - \frac{\partial f}{\partial v_i}\frac{\partial g}{\partial x_i} \ , \tag{1.35}$$

where n is the number of degrees of freedom. It can be proven (see Whittaker, 1937, Section 32, or Gantmacher, 1975, Section 131) that a transformation $(\mathbf{v}, \mathbf{x}) \to (\mathbf{v}', \mathbf{x}')$ is canonical if and only if, considering the components $v'_1, \ldots, v'_n$ of $\mathbf{v}'$ and $x'_1, \ldots, x'_n$ of $\mathbf{x}'$ as functions of $(\mathbf{v}, \mathbf{x})$, one has

$$\{v'_i, v'_j\} = 0 \ , \quad \{x'_i, x'_j\} = 0 \ , \quad \{x'_i, v'_j\} = \delta_{i,j} \ , \tag{1.36}$$

where $\delta_{i,j}$ is 1 if $i = j$ and 0 otherwise.

This is a very practical and useful criterion to check whether a given transformation is canonical, and we will often use it throughout this book. As a first example, let's go back to the Hamiltonian (1.32) to see how heliocentric coordinates could be introduced in a canonical way. Defining the new coordinates $\mathbf{r}_j = \mathbf{u}_j - \mathbf{u}_0$ for $j = 1, \ldots, N$ (the heliocentric positions of the planets) and $\mathbf{r}_0 = \mathbf{u}_0$, the problem is to find new momenta $\mathbf{v}_j$ such that the transformation $(\tilde{\mathbf{u}}_j, \mathbf{u}_j) \to (\mathbf{v}_j, \mathbf{r}_j)$ is canonical. The application of criterion (1.36) tells us that the good transformation is $\mathbf{v}_0 = \tilde{\mathbf{u}}_0 + \tilde{\mathbf{u}}_1 + \cdots + \tilde{\mathbf{u}}_N$ and $\mathbf{v}_j = \tilde{\mathbf{u}}_j$ for $j = 1, \ldots, N$. With this choice of new canonical coordinates and momenta, the new Hamiltonian becomes, as it is easy to check by substitution of the new variables in (1.32)

$$\begin{aligned}\mathcal{H} = {} & \frac{\|\mathbf{v}_0\|^2}{2m_0} - \sum_{j=1}^{N} \frac{\mathbf{v}_0 \cdot \mathbf{v}_j}{m_0} + \frac{1}{2} \sum_{j=1}^{N} \|\mathbf{v}_j\|^2 \left[\frac{1}{m_j} + \frac{1}{m_0}\right] + \sum_{j=1}^{N} \sum_{i=1}^{j-1} \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{m_0} \\ & - \mathcal{G} \sum_{j=1}^{N} \sum_{i=1}^{j-1} \frac{m_i m_j}{\|\boldsymbol{\Delta}_{i,j}\|} - \mathcal{G} \sum_{j=1}^{N} \frac{m_0 m_j}{\|\mathbf{r}_j\|} \ ,\end{aligned} \tag{1.37}$$

with $\boldsymbol{\Delta}_{i,j} = \mathbf{r}_i - \mathbf{r}_j$. Since the Hamiltonian does not depend on $\mathbf{r}_0$, the equations of motion give $\dot{\mathbf{v}}_0 = 0$ (which means that the baricenter of the system keeps its inertial motion). Without loss of generality we can then assume that $\mathbf{v}_0 = 0$ and drop the terms in $\mathbf{v}_0$ from the Hamiltonian. The heliocentric Hamiltonian of a planetary system then reads

$$\mathcal{H} = \sum_{j=1}^{N} \left\{ \frac{\|\mathbf{v}_j\|^2 (m_0 + m_j)}{2 m_0 m_j} - \mathcal{G} \frac{m_0 m_j}{\|\mathbf{r}_j\|} \right\} + \sum_{j=1}^{N} \sum_{i=1}^{j-1} \left\{ \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{m_0} - \mathcal{G} \frac{m_i m_j}{\|\boldsymbol{\Delta}_{i,j}\|} \right\} \ . \tag{1.38}$$

Note that the Hamiltonian $\mathcal{H}$ can be read as the sum of Hamiltonians

$$\mathcal{H}_0^j = \frac{\|\mathbf{v}_j\|^2}{2\mu_j} - \frac{\mathcal{G}(m_0 + m_j)\mu_j}{\|\mathbf{r}_j\|} \tag{1.39}$$

of two-body problems of type (1.26) with reduced masses $\mu_j = m_0 m_j/(m_0 + m_j)$, perturbed by terms whose size relative to the two-body problem Hamiltonians is proportional to the mass of the planets relative to that of the Sun.

Rearranging the kinetic terms, (1.38) can be written as

$$\mathcal{H} = \sum_{j=1}^{N} \left\{ \frac{\|\mathbf{v}_j\|^2}{2m_j} - \mathcal{G}\frac{m_0 m_j}{\|\mathbf{r}_j\|} \right\} + \frac{1}{2m_0} \left\| \sum_{j=1}^{N} \mathbf{v}_j \right\|^2 - \mathcal{G} \sum_{j=1}^{N} \sum_{i=1}^{j-1} \frac{m_i m_j}{\|\mathbf{\Delta}_{i,j}\|} \; . \tag{1.40}$$

This formulation of the Hamiltonian (sometimes called *democratic heliocentric*) is of particular interest for the development of algorithms of symplectic integration (Koseleff, 1993, 1996; Touma and Wisdom, 1994b; Duncan *et al.*, 1998; Chambers, 1999).

After this digression, let's now come back to the criteria to discriminate canonical transformations from all other possible transformation on the phase space.

A second criterion, which is not very easy to handle in practical computations, but will be used later in the Arnold–Liouville theorem (see Section 1.9) is the following: the transformation $(\mathbf{v}, \mathbf{x}) \to (\mathbf{v}', \mathbf{x}')$ is canonical if there exists a function $S(\mathbf{v}', \mathbf{x})$ (called a *generating function*) such that

$$v_j = \frac{\partial S}{\partial x_j}(\mathbf{v}', \mathbf{x}) \; , \quad x'_j = \frac{\partial S}{\partial v'_j}(\mathbf{v}', \mathbf{x}) \; , \quad 1 \le j \le n \tag{1.41}$$

(see Chapter 4 of Whittaker, 1937). Note that not all canonical transformations can be written as in (1.41).

A third criterion, which is at the basis of Lie's approach to perturbation theory – as will be illustrated in Chapter 2 – is: the transformation $(\mathbf{v}, \mathbf{x}) \to (\mathbf{v}', \mathbf{x}')$ is canonical if there exists a Hamiltonian $\chi(\mathbf{v}', \mathbf{x}')$ (called a *generating Hamiltonian*) and a parameter ε such that

$$\mathbf{v} = \mathbf{v}' + \int_0^{\varepsilon} \dot{\mathbf{v}}' \mathrm{d}t \equiv \mathbf{v}'(\varepsilon) \; , \quad \mathbf{x} = \mathbf{x}' + \int_0^{\varepsilon} \dot{\mathbf{x}}' \mathrm{d}t \equiv \mathbf{x}'(\varepsilon) \; , \tag{1.42}$$

where $\dot{\mathbf{v}}'$ and $\dot{\mathbf{x}}'$ are given by the Hamiltonian equations $\dot{\mathbf{v}}' = -\partial\chi/\partial\mathbf{x}'$ and $\dot{\mathbf{x}}' = \partial\chi/\partial\mathbf{v}'$ (see Gantmacher, 1975, Section 133-134 for a proof). In other words, a transformation on the phase space is canonical if it can be interpreted as the result of a Hamiltonian flow at a certain time ε. For instance, a rotation in phase space by an angle ϑ in one degree of freedom, namely $x = x'\cos\vartheta + v'\sin\vartheta$, $v = -x'\sin\vartheta + v'\cos\vartheta$ is canonical because it can be interpreted as the flow at time $t = \vartheta$ generated by the Hamiltonian $\chi(v', x') = 1/2(v'^2 + x'^2)$ of the harmonic oscillator. Not all canonical transformations can be written as the result of a Hamiltonian flow (a counterexample is given by the canonical transformation from Cartesian to polar coordinates $x = \sqrt{2p}\cos q, y = \sqrt{2p}\sin q$).

1.7 Properties of Hamiltonian flow

By *Hamiltonian flow* we mean the time evolution of coordinates and momenta described by the Hamiltonian equations. Hamiltonian flow has some important specific properties, that will be useful in the following.

A) *Conservation of volume (Liouville's theorem).* As is well known, for a system of equations of type (1.19), the volume δV of an infinitesimal set of initial conditions evolves with time according to the equation

$$\frac{1}{\delta V}\frac{\mathrm{d}\delta V}{\mathrm{d}t} = \sum_{j=1}^{m} \frac{\partial F_i}{\partial r_i} \equiv \mathrm{div}\mathbf{F} \ , \tag{1.43}$$

where m is the number of components of $\mathbf{r}$ and $\mathbf{F}$. For a Hamiltonian system (1.20), where $m = 2n$, $\mathbf{r} = (v_1, \ldots, v_n, x_1, \ldots, x_n)$ and

$$\mathbf{F} = \left(-\frac{\partial \mathcal{H}}{\partial x_1}, \ldots, -\frac{\partial \mathcal{H}}{\partial x_n}, \frac{\partial \mathcal{H}}{\partial v_1}, \ldots, \frac{\partial \mathcal{H}}{\partial v_n} \right) \tag{1.44}$$

it is trivial to check that $\mathrm{div}\mathbf{F} = 0$. This means that the volume is preserved by the Hamiltonian flow. This property is very important, because it implies that a cloud of initial conditions can never shrink nor inflate. In particular, it also implies that Hamiltonian dynamics cannot have *attractors*, namely manifolds of dimension smaller than the number of degrees of freedom towards which the flow may collapse. This will be very important for the considerations that will be covered in Chapter 4.

B) *Conservation of* $\mathcal{H}$. The rate by which $\mathcal{H}$ changes with time can be written, by differentiation of $\mathcal{H}(\mathbf{v}, \mathbf{x}, t)$, as

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \mathrm{grad}_{\mathbf{x}}\mathcal{H} \cdot \dot{\mathbf{x}} + \mathrm{grad}_{\mathbf{v}}\mathcal{H} \cdot \dot{\mathbf{v}} + \frac{\partial \mathcal{H}}{\partial t} \ . \tag{1.45}$$

Using the Hamiltonian equations (1.20) for $\dot{\mathbf{v}}$ and $\dot{\mathbf{x}}$ one gets

$$\frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} = \frac{\partial \mathcal{H}}{\partial t} \ . \tag{1.46}$$

In other words, the Hamiltonians which do not depend explicitly on time (i.e. the autonomous Hamiltonians) do not change value along the flow that they generate, namely they are *constants of motion.* For the systems (1.23) such that $T_1 = T_0 = 0$, $\mathcal{H}$ is the total energy of the system, so that the conservation of the value of the Hamiltonian is nothing but the conservation of energy. However, this is not true in general, when T_1 or T_0 are not null.

C) *Time evolution of a function along a Hamiltonian flow.* Given a function $f(\mathbf{v}, \mathbf{x})$ defined on phase space, with $\mathbf{v}$ and $\mathbf{x}$ evolving according to the Hamiltonian equations (1.20), one gets by differentiation

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \mathrm{grad}_{\mathbf{x}} f \cdot \dot{\mathbf{x}} + \mathrm{grad}_{\mathbf{v}} f \cdot \dot{\mathbf{v}} = \{f, \mathcal{H}\} \ . \tag{1.47}$$

The value $f(t) \equiv f(\mathbf{v}(t), \mathbf{x}(t))$ is called the *evolution* of f along the flow of $\mathcal{H}$. Formula (1.47) allows one to write $f(t)$ as a function of $t, \mathbf{v}(0)$ and $\mathbf{x}(0)$ as follows. For t small enough, one expands $f(t)$ in a Taylor series as

$$f(t) = f(0) + \sum_{i=1}^{\infty} \frac{t^i}{i!} \frac{\mathrm{d}^i f}{\mathrm{d}t^i}(0), \tag{1.48}$$

where $f(0) \equiv f(\mathbf{v}(0), \mathbf{x}(0))$ and $\mathrm{d}^i f/\mathrm{d}t^i(0) \equiv \mathrm{d}^i f/\mathrm{d}t^i(\mathbf{v}(0), \mathbf{x}(0))$, and uses the relations

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \{f, \mathcal{H}\} \ , \quad \frac{\mathrm{d}^2 f}{\mathrm{d}t^2} = \left\{\frac{\mathrm{d}f}{\mathrm{d}t}, \mathcal{H}\right\} = \{\{f, \mathcal{H}\}, \mathcal{H}\} \ , \quad \ldots \tag{1.49}$$

Then, denoting

$$\mathcal{L}^1_{\mathcal{H}} f = \{f, \mathcal{H}\} \ , \quad \mathcal{L}^i_{\mathcal{H}} f = \mathcal{L}^1_{\mathcal{H}} \mathcal{L}^{i-1}_{\mathcal{H}} f \ , \tag{1.50}$$

one finally gets

$$f(t) = f(0) + \sum_{i=1}^{\infty} \frac{t^i}{i!} \mathcal{L}^i_{\mathcal{H}} f(0) \ . \tag{1.51}$$

This expansion is usually called the *Lie series* of f under the flow of $\mathcal{H}$, and we will denote it hereafter by $S^t_{\mathcal{H}} f$. As anticipated above it allows one to write $f(\mathbf{v}(t), \mathbf{x}(t))$ as a function of $\mathbf{v}(0), \mathbf{x}(0)$, the time t playing the role of a parameter, and it will become very important in Chapter 2 to design a suitable perturbation approach to study Hamiltonian systems.

Finally, using (1.51), the canonical transformation (1.42) can be rewritten

$$\mathbf{v} = S^{\varepsilon}_{\chi} \mathbf{v}' \ , \quad \mathbf{x} = S^{\varepsilon}_{\chi} \mathbf{x}' \ , \tag{1.52}$$

where the vectorial notation $\boldsymbol{\alpha} = S^{\varepsilon}_{\chi} \boldsymbol{\alpha}'$ denotes $\alpha_1 = S^{\varepsilon}_{\chi} \alpha'_1, \ldots, \alpha_n = S^{\varepsilon}_{\chi} \alpha'_n$.

1.8 Integrable Hamiltonians

The solution of the system of differential equations

$$\frac{\mathrm{d}r_i}{\mathrm{d}t} = F_i(\mathbf{r}) \ , \quad \text{with } i = 1, \ldots, n \quad \text{and } \mathbf{r} \equiv (r_1, \ldots, r_n) \tag{1.53}$$

can be written in implicit form as a system of integral equations

$$\int_{\mathbf{r}(0)}^{\mathbf{r}(t)} \frac{\mathrm{d}r_i}{F_i(\mathbf{r})} = \int_0^t \mathrm{d}t \ . \tag{1.54}$$

The system (1.53) is therefore said to be *integrable*, if the integrals on the left-hand side of (1.54) can be explicitly computed, and the resulting relationships $F_i'(\mathbf{r}(t)) - F_i'(\mathbf{r}(0)) = t$, where F_i' are the primitives of $1/F_i$, can be inverted, giving $\mathbf{r}(t)$ as an explicit function of t (see chapter 4 of Arnold *et al.*, 1988).

For example, the differential equation $\mathrm{d}x/\mathrm{d}t = x$ is integrable, because the integral equation $\int_{x(0)}^{x(t)} \mathrm{d}x/x = \int_0^t \mathrm{d}t$ gives $\ln x(t) - \ln x(0) = t$, namely $x(t) = x(0)\exp(t)$.

Using this definition of integrability, it is very difficult to conclude whether a given system of differential equations is integrable or not. If the solutions of the integrals are not found, it is hard to know if this is due to a genuine lack of integrability of the system, or simply to a lack of skill in finding the primitives F_i'.

For Hamiltonian systems, partial help comes from *Liouville's theorem*, which states that an n-degree of freedom Hamiltonian is integrable if it admits n independent constants of motion $\Phi_1, \ldots, \Phi_n$, such that $\{\Phi_i, \Phi_j\} = 0$ for $i \neq j$. Although it is easier to find constants of motion than to actually solve the Hamiltonian equations, there is no general recipe on how *all* constants of motion can be found. In particular, if only m constants of motion are known, with $m < n$, it is hard to know if additional constants of motion are still to be found or really don't exist. For instance, celestial mechanicians looked for several years for a third constant of motion of the three-degree of freedom Hamiltonian that describes the motion of a star in a cubic galactic potential, until Hénon and Heiles (1964) numerically showed that such a third constant of motion does not exist. Conversely, the Toda lattice Hamiltonian has been long conjectured to be nonintegrable, until M. Hénon (1974) found the last missing constant of motion!

Luckily nowadays the situation is not so desperate as it was before the work of Poincaré (1892). There exists a criterion of nonintegrability – that is the appearance of chaos, which will be discussed in Chapter 4 – that can be used both analytically and numerically. In this book, we will often be led to introduce integrable approximations of the real dynamics. These integrable Hamiltonians will be of only three kinds:

i) They will depend only on the momenta of the system, i.e. $\mathcal{H}(v_1, \ldots, v_n)$. In this case the solution is trivial. The momenta v_i are constants of motion (because $\partial\mathcal{H}/\partial x_i = 0$), and the coordinates x_i move linearly with time, with constant speeds $\omega_i = \partial\mathcal{H}/\partial v_i$, for $1 \leq i \leq n$.

ii) They will have only one degree of freedom, i.e. $\mathcal{H}(v, x)$. In this case the Hamiltonian system is integrable because it has one constant of motion, that is the Hamiltonian itself. The motion will then evolve along level curves of $\mathcal{H}$ on the two-dimensional phase space (v, x).

iii) They will depend only on one coordinate, i.e. $\mathcal{H}(v_1, \ldots, v_n, x_k)$. In this case the system is integrable because it has n independent constants of motion which are $v_1, \ldots, v_{k-1}$, $v_{k+1}, \ldots, v_n$ and $\mathcal{H}$. The motion evolves preserving the value of $v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n$, and follows level curves of $\mathcal{H}$ on the plane (v_k, x_k).

1.9 Action–angle variables

For integrable Hamiltonian systems, of crucial importance is the *Arnold–Liouville theorem*, an extension of Liouville's theorem (see Section 1.8) found by Arnold (1963a). Arnold proved that, in the hypotheses of Liouville's theorem and if the n-dimensional surface implicitly defined by the constants of motion $\Phi_1, \ldots, \Phi_n$ is *compact*, it is then possible to introduce canonical momenta $\mathbf{p}$ and coordinates $\mathbf{q}$ such that

i) The coordinates $q_1, \ldots, q_n$ are *angles*, cyclically defined on the interval $[0, 2\pi]$, and the canonical transformation from the original momenta and coordinates, i.e. $\mathbf{v}(\mathbf{p}, \mathbf{q}), \mathbf{x}(\mathbf{p}, \mathbf{q})$, is 2π-periodic on the angles $q_1, \ldots, q_n$.

ii) In the new variables, the Hamiltonian is a function of the momenta $\mathbf{p}$ only, i.e. $\mathcal{H} \equiv \mathcal{H}(\mathbf{p})$.

The momenta $\mathbf{p}$ are usually called the *actions* of the system. A set of canonical variables $(\mathbf{p}, \mathbf{q})$, where the coordinates $\mathbf{q}$ are angles will be generically called *action–angle variables*. Although action–angle variables had been previously used by several authors (Epstein, 1916; Sommerfeld, 1922; Born, 1927) on specific problems, the Arnold–Liouville theorem is very important because it shows that basically any integrable Hamiltonian can be written, in suitable action–angle variables, as a function of the sole actions. Therefore, in the light of the Arnold–Liouville theorem, one can generically represent the integrable Hamiltonians by functions $\mathcal{H}(\mathbf{p})$, and work out the general theory of quasi-integrable Hamiltonian dynamics in action–angle variables – as will be done hereafter in this book.

Moreover, the proof of the Arnold–Liouville theorem also provides a constructive recipe for the introduction of action–angle variables in practical cases.

The existence of n constants of motion for an n-degree of freedom Hamiltonian system ensures that the motion evolves on an n-dimensional surface $M_{\mathbf{\Phi}}$ embedded in $2n$-dimensional phase space. The fact that $\{\Phi_i, \Phi_j\} = 0$ for $i \neq j$ ensures that the motion can be decomposed in n independent flows, generated by the functions $\Phi_1, \ldots, \Phi_n$, each considered as a one-degree of freedom Hamiltonian. This means that the evolution of the motion at time t, i.e. $\mathbf{v}(t), \mathbf{x}(t)$, can be obtained following the flow of Φ_1 for a time t, from the initial condition $\mathbf{v}(0), \mathbf{x}(0)$ to a point $\mathbf{v}_1, \mathbf{x}_1$, then following the flow of Φ_2 for a time t, from $\mathbf{v}_1, \mathbf{x}_1$ to another point $\mathbf{v}_2, \mathbf{x}_2$, and so on. The final point $\mathbf{v}_n, \mathbf{x}_n$ will coincide with $\mathbf{v}(t), \mathbf{x}(t)$. The condition that the surface $M_{\mathbf{\Phi}}$ is compact implies that the individual flows of $\Phi_1, \ldots, \Phi_n$, and hence the global motion, can be decomposed into independent periodic cycles, which we denote by $\gamma_1, \ldots, \gamma_n$. The actions $\mathbf{p}$ are then introduced by

$$p_i = \frac{1}{2\pi} \oint_{\gamma_i} \sum_{j=1}^{n} v_j \mathrm{d}x_j \ . \tag{1.55}$$

Then, writing $\mathbf{v}$ as functions of $\mathbf{p}$ and $\mathbf{x}$, the integral generating function

$$S(\mathbf{p}, \mathbf{x}) \equiv \int \sum_{j=1}^{n} v_j(\mathbf{p}, \mathbf{x}) \mathrm{d}x_j \tag{1.56}$$

is defined, and the new coordinates $\mathbf{q}$ are introduced as

$$q_i = \frac{\partial S}{\partial p_i}(\mathbf{p}, \mathbf{x}) \ . \tag{1.57}$$

The transformation $(\mathbf{v}, \mathbf{x}) \to (\mathbf{p}, \mathbf{q})$ so defined is of the form (1.41) and therefore is, by construction, canonical. One can prove that $q_1, \ldots, q_n$ are angles, namely q_i is increased by 2π when a complete cycle γ_i is followed, and that the Hamiltonian $\mathcal{H}$ is dependent on the actions $\mathbf{p}$ only (see Arnold, 1963a).

1.9.1 Delaunay variables

As an example of the application of the Arnold–Liouville theorem, let's proceed to introduce the action–angle variables for the integrable Hamiltonian of the two-body problem. They will be the variables that we will later use to study the dynamics of the restricted problem and of the planetary problem using Hamiltonian perturbation techniques. We follow here the approach of Born (1927), elaborated for the equivalent problem of classical motion of an electron around the core of the hydrogen atom.

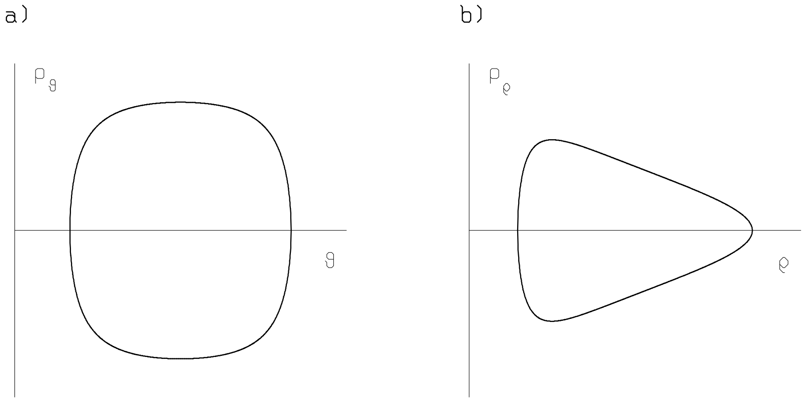


Figure 1.3: (a) The cycle γ_ϑ defined by the constant of motion G^2; (b) the cycle γ_ϱ defined by the constant of motion $\mathcal{H}$.

It is convenient to start from the two-body problem Hamiltonian in spherical coordinates (1.24). We first identify three independent constants of motion:

$$H = p_\varphi \ , \quad G^2 = p_\vartheta^2 + \frac{H^2}{\sin^2\vartheta} \ , \quad \mathcal{H} = \frac{1}{2}\left(p_\varrho^2 + \frac{G^2}{\varrho^2}\right) - \frac{\mathcal{G}(m_0+m_1)}{\varrho} \ , \tag{1.58}$$

and check that the conditions $\{H, G^2\} = 0$, $\{H, \mathcal{H}\} = 0$ and $\{G^2, \mathcal{H}\} = 0$ trivially hold (the last two Poisson brackets also prove that H and G^2 are constants of motion for $\mathcal{H}$). Using the definition of the momenta p_ϑ and p_φ, it is easy to see that G is the norm of the angular momentum of the system, while H is the component of the angular momentum vector along the z axis, both normalized by the reduced mass. The constant $\mathcal{H}$ is the Hamiltonian of the two-body problem.

Then we proceed to identify the cycles. The Keplerian case is easier than the general case, because each of the constants of motion (1.58) straightforwardly defines a cycle. The constant H defines a cycle γ_φ on the (p_φ, φ) plane, which is given by $p_\varphi = H$ and $\varphi \in [0, 2\pi]$ (remember that φ is already defined as an angle). The constant G^2 defines a less trivial cycle γ_ϑ, shown in Fig. 1.3a, on the (p_ϑ, ϑ) plane. Along this cycle, ϑ oscillates between two values $\vartheta_{\min}$

and $\vartheta_{\max}$, with $\vartheta_{\min}$, $\vartheta_{\max}$ and $p_\vartheta(\vartheta)$ given by the relations

$$\vartheta_{\min} = \arcsin\left|\frac{H}{G}\right| \ , \quad \vartheta_{\max} = 2\pi - \vartheta_{\min} \ , \quad p_\vartheta(\vartheta) = \pm\sqrt{G^2 - \frac{H^2}{\sin^2\vartheta}} \ . \quad (1.59)$$

Finally the constant $\mathcal{H}$ defines – if it is negative – a cycle γ_ϱ in the (p_ϱ, ϱ) plane, illustrated in Fig.1.3b. The condition that $\mathcal{H}$ is negative is nothing but the condition that the surface M_Φ is compact, which is crucial in the Arnold–Liouville theorem. If $\mathcal{H}$ were not negative, the motion would be unbounded and γ_ϱ would not be a closed cycle. Along this cycle, ϱ oscillates between two positive values $\varrho_{\min}$ and $\varrho_{\max}$, with

$$\begin{aligned}
\varrho_{\min} &= \{-\mathcal{G}(m_0+m_1) + \sqrt{\mathcal{G}^2(m_0+m_1)^2 + 2\mathcal{H}G^2}\}/2\mathcal{H} \ , \\
\varrho_{\max} &= \{-\mathcal{G}(m_0+m_1) - \sqrt{\mathcal{G}^2(m_0+m_1)^2 + 2\mathcal{H}G^2}\}/2\mathcal{H} \ , \\
p_\varrho(\varrho) &= \pm\sqrt{2\,(\mathcal{H} + \mathcal{G}(m_0+m_1)/\varrho) - G^2/\varrho^2} \ . \qquad (1.60)
\end{aligned}$$

The following step is to introduce the actions of the system, using (1.55). Because on the cycle γ_φ only φ evolves, $d\vartheta$ and $d\varrho$ are zero, and the sum in (1.55) reduces to the sole term $p_\varphi d\varphi$; an analogous situation happens for the cycles γ_ϑ and γ_ϱ. Therefore, using (1.59) and (1.60) the actions become:

$$\begin{aligned}
p_1 &= \frac{1}{2\pi}\oint_{\gamma_\varphi} p_\varphi \mathrm{d}\varphi = \frac{1}{2\pi}\int_0^{2\pi} H\mathrm{d}\varphi = H \\
p_2 &= \frac{1}{2\pi}\oint_{\gamma_\vartheta} p_\vartheta \mathrm{d}\vartheta = \frac{1}{\pi}\int_{\vartheta_{\min}}^{\vartheta_{\max}} p_\vartheta(\vartheta)\mathrm{d}\vartheta = G - H \\
p_3 &= \frac{1}{2\pi}\oint_{\gamma_\varrho} p_\varrho \mathrm{d}\varrho = \frac{1}{\pi}\int_{\varrho_{\min}}^{\varrho_{\max}} p_\varrho(\varrho)\mathrm{d}\varrho = -G + \sqrt{-\frac{\mathcal{G}^2(m_0+m_1)^2}{2\mathcal{H}}} \ . \quad (1.61)
\end{aligned}$$

Using these actions, it is easy to compute, by inversion of (1.61), that the Hamiltonian of the two-body problem is

$$\mathcal{H} = -\frac{\mathcal{G}^2(m_0+m_1)^2}{2(p_1+p_2+p_3)^2} \ . \qquad (1.62)$$

Finally, to introduce the conjugate angles q_1, q_2, q_3, we first define the generating function

$$S(p_1,p_2,p_3,\varphi,\vartheta,\varrho) = \int p_\varrho(p_1,p_2,p_3,\varrho)\mathrm{d}\varrho + p_\vartheta(p_1,p_2,\vartheta)\mathrm{d}\vartheta + p_\varphi(p_1)\mathrm{d}\varphi \quad (1.63)$$

where the expressions of $p_\varrho(p_1,p_2,p_3,\varrho)$, $p_\vartheta(p_1,p_2,\vartheta)$ and $p_\varphi(p_1)$ are obtained by inverting (1.58) and using the relations (1.61). The conjugate angles will then be:

$$q_1 = \frac{\partial S}{\partial p_1} \ , \quad q_2 = \frac{\partial S}{\partial p_2} \ , \quad q_3 = \frac{\partial S}{\partial p_3} \ . \qquad (1.64)$$

The construction of canonical action–angle variables for the two-body problem is therefore complete.

Once a set of action–angle variables is introduced, any linear transformation of the type

$$\mathbf{J} = A\mathbf{p} \ , \quad \boldsymbol{\psi} = (A^T)^{-1}\mathbf{q} \ , \tag{1.65}$$

where A is a matrix with integer coefficients and determinant equal to 1, and $(A^T)^{-1}$ is the inverse of its transposition, defines new canonical action–angle variables. Moreover, if the Hamiltonian depends of the actions $\mathbf{p}$ only, then it will obviously depend only on the actions $\mathbf{J}$. In other words, Arnold action–angle variables are not unique. In the case of the two-body problem, because of the form of (1.62), it is convenient to use as new canonical action–angle variables the following

$$\begin{aligned} L &= p_1 + p_2 + p_3 \ , & l &= q_3 \\ G &= p_1 + p_2 \ , & g &= q_2 - q_3 \\ H &= p_1 \ , & h &= q_1 - q_2 \ . \end{aligned} \tag{1.66}$$

With this choice, the Hamiltonian of the two-body problem will be simply

$$\mathcal{H} = -\frac{\mathcal{G}^2 (m_0 + m_1)^2}{2L^2} \ . \tag{1.67}$$

The canonical action–angle variables L, l, G, g, H, h are usually called the *Delaunay variables.* The relationship of the actions G and H to the orbital elements defined in Section 1.2 is already given in formulæ (1.14) and (1.15) respectively, while the expression for L can be easily obtained from (1.13). Conversely, in order to establish the relationship of the angles l, g and h with the orbital elements it is necessary to compute the integral equations (1.64) and use (1.9). The resulting relationships are:

$$\begin{aligned} L &= \sqrt{\mathcal{G}(m_0 + m_1)a} \ , & l &= M \\ G &= L\sqrt{1 - e^2} \ , & g &= \omega \\ H &= G \cos i \ , & h &= \Omega \ . \end{aligned} \tag{1.68}$$

To avoid the problem that the angles l, g, h are not well defined when the inclination and/or the eccentricity are zero, the following *modified Delaunay*

variables are also often used:

$$\begin{aligned} \Lambda &= L = \sqrt{\mathcal{G}(m_0+m_1)a} \ , & \lambda &= l+g+h = M+\varpi \\ P &= L-G = L(1-\sqrt{1-e^2}) \ , & p &= -g-h = -\varpi \\ Q &= G-H = 2G\sin^2\tfrac{i}{2} \ , & q &= -h = -\Omega \ . \end{aligned} \tag{1.69}$$

These variables have the advantage that λ is always well defined, while p and q are not defined only when the conjugate actions P and Q are respectively equal to zero. Therefore, (P, p) and (Q, q) constitute polar coordinate systems. Note that, for small eccentricities and inclinations, P is proportional to e^2 and Q to i^2.

1.9.2 Hamilton equations in Delaunay variables for the restricted and the planetary problem

The Hamiltonian of the restricted problem (1.29) can be rewritten in Delaunay's variables as

$$\mathcal{H} = -\frac{\mathcal{G}^2 m_0^2}{2L^2} + \mathcal{H}_1(L, G, H, l, g, h, t) \ , \tag{1.70}$$

where $\mathcal{H}_1(L, G, H, l, g, h, t)$ is obtained by rewriting

$$-\mathcal{G}\sum_{j=1}^{N} m_j \left(\frac{1}{\|\mathbf{\Delta}_j\|} - \frac{\mathbf{r}\cdot\mathbf{s}_j}{\|\mathbf{s}_j\|^3} \right) \ , \tag{1.71}$$

in osculating elements using (1.9) and the inverse of (1.68). The time t enters through the position vectors of the planets $\mathbf{s}_j$. Although the resulting expression is quite long, it can be explicitly written and used for practical computations. The Hamiltonian equations are

$$\begin{aligned} \dot{L} &= -\frac{\partial \mathcal{H}_1}{\partial l} \ , & \dot{l} &= \frac{\mathcal{G}^2 m_0^2}{L^3} + \frac{\partial \mathcal{H}_1}{\partial L} \\ \dot{G} &= -\frac{\partial \mathcal{H}_1}{\partial g} \ , & \dot{g} &= \frac{\partial \mathcal{H}_1}{\partial G} \\ \dot{L} &= -\frac{\partial \mathcal{H}_1}{\partial h} \ , & \dot{h} &= \frac{\partial \mathcal{H}_1}{\partial H} \ . \end{aligned} \tag{1.72}$$

For the Hamiltonian of the planetary problem (1.38) the introduction of the Delaunay variables is a little more complicated. First notice that the integrable approximation of (1.38) is the sum of Hamiltonians (1.39) which are

two-body-problem Hamiltonians, but in the form (1.26) rather than (1.22). Had we started the construction of action–angle variables from (1.26), the resulting actions would have been multiplied by μ_1 with respect to those defined in (1.68), and the resulting Hamiltonian would have been multiplied by μ_1^3 with respect to (1.67).[3] Second, recall that the momenta $\mathbf{v}_j$ in (1.39) are *not* equal to $\mu_j \dot{\mathbf{r}}_j$. This however does not change the relationships between $\mathbf{v}_j, \mathbf{r}_j$ and the Delaunay variables $L_j, G_j, H_j, l_j, g_j, h_j$ with respect to those valid for a classical two-body problem where $\mathbf{v}_j = \mu_j \dot{\mathbf{r}}_j$; in fact the Hamiltonians (1.39) and (1.26) are formally identical, *regardless* of what the momenta really represent. For this reason, we can define *formal* osculating elements $a_j, e_j, i_j, M_j, \omega_j, \Omega_j$ by using formula (1.9) with $\mathbf{v}_j/\mu_j$ instead of $\mathrm{d}\mathbf{r}/\mathrm{d}t$, and write the Delaunay variables for the planetary problem as

$$\begin{aligned} L_j &= \mu_j \sqrt{\mathcal{G}(m_0 + m_j) a_j}\ , & l_j &= M_j \\ G_j &= L_j \sqrt{1 - e_j^2}\ , & g_j &= \omega_j \\ H_j &= G_j \cos i_j\ , & h_j &= \Omega_j\ . \end{aligned} \tag{1.73}$$

The Hamiltonian (1.38) then becomes

$$\mathcal{H} = \sum_j -\frac{\mathcal{G}^2 (m_0 + m_j)^2 (\mu_j)^3}{2L_j^2} + \mathcal{H}_1 \tag{1.74}$$

where $\mathcal{H}_1$ can be explicitly written as a function of the Delaunay variables (1.73) by direct substitution in (1.38), similarly to what is done in the case of the restricted problem. The Hamiltonian equation will then be, for each j:

$$\begin{aligned} \dot{L}_j &= -\frac{\partial \mathcal{H}_1}{\partial l_j}\ , & \dot{l}_j &= \frac{\mathcal{G}^2 (m_0 + m_j)^2 (\mu_j)^3}{L_j^3} + \frac{\partial \mathcal{H}_1}{\partial L_j} \\ \dot{G}_j &= -\frac{\partial \mathcal{H}_1}{\partial g_j}\ , & \dot{g}_j &= \frac{\partial \mathcal{H}_1}{\partial G_j} \\ \dot{L}_j &= -\frac{\partial \mathcal{H}_1}{\partial h_j}\ , & \dot{h}_j &= \frac{\partial \mathcal{H}_1}{\partial H_j}\ . \end{aligned} \tag{1.75}$$

Equations (1.72) and (1.75) are the Hamiltonian version of the Lagrange equations.

[3]This can be easily seen as follows. Multiply (1.26) by μ_1. Then (1.26) becomes formally equal to (1.22), with constant $\mathcal{G}' = \mathcal{G}\mu_1^2$ replacing $\mathcal{G}$. The Delaunay variables and the resulting Hamiltonian will then be as in (1.68) and (1.67), with $\mathcal{G}'$ instead of $\mathcal{G}$. Finally divide the resulting Hamiltonian by μ_1.

As for the restricted problem, to avoid the apparent singularities occurring in (1.75) when some inclination and/or the eccentricity is zero, one usually uses the *modified Delaunay variables*:

$$\begin{aligned} \Lambda_j &= L_j \ , & \lambda_j &= l_j + g_j + h_j \\ P_j &= L_j - G_j \ , & p_j &= -g_j - h_j \\ Q_j &= G_j - H_j \ , & q_j &= -h_j \ . \end{aligned} \tag{1.76}$$

1.9.3 D'Alembert rules

In classical studies of the dynamics of the restricted and planetary problems, the function $\mathcal{H}_1$ in (1.70) and (1.74) is usually expanded in Fourier series of the angles λ, ϖ and Ω and in power series of e and i, or equivalently, using the modified Delaunay variables (1.69), in Fourier series of λ, p, q and in power series of $P^{1/2}, Q^{1/2}$. Although in this book we will never deal with classical explicit expansions, in the following chapters it will be very important to know their general form.

Let us denote by $\Lambda_j, P_j, Q_j, \lambda_j, p_j, q_j$ $(j = 1, \dots, N)$ the modified Delaunay variables of N bodies (one small body and N planets, or N planets) and by $\alpha_j, \beta_j, k_j, m_j, s_j$ integer numbers. Moreover we generically denote by $\mathbf{\Lambda}$, $\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{k}, \mathbf{m}$ and $\mathbf{s}$ the vectors whose components are Λ_j, $\alpha_j, \beta_j, k_j, m_j$ and s_j respectively. The most general form of the Fourier series expansion in the angles and power series expansion in $\sqrt{P_j}, \sqrt{Q_j}$ is therefore:

$$\mathcal{H}_1 = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{k},\mathbf{m},\mathbf{s}} c_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{k},\mathbf{m},\mathbf{s}}(\mathbf{\Lambda}) \left(\prod_j P_j^{\alpha_j/2} Q_j^{\beta_j/2} \right) \exp\left[\iota\left(\sum_j k_j\lambda_j + \sum_j m_j p_j + \sum_j s_j q_j\right)\right] \ , \tag{1.77}$$

where ι denotes $\sqrt{-1}$, and $c_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{k},\mathbf{m},\mathbf{s}}$ are suitable coefficients (for their computation see Brouwer and Clemence, 1961; Duriez, 1989; Laskar and Robutel, 1995; Ellis and Murray, 2000).

Consideration of the symmetries and analytic properties of $\mathcal{H}_1$ allow the easily derivation of the so-called *D'Alembert rules*:

1) $\mathcal{H}_1$ must be invariant under a simultaneous change of sign of all the angles λ_j, p_j, q_j. Therefore the Fourier series expansion must contain only cosine terms, namely $c_{\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{k},\mathbf{m},\mathbf{s}} = c_{\boldsymbol{\alpha},\boldsymbol{\beta},-\mathbf{k},-\mathbf{m},-\mathbf{s}}$ and all coefficients are real.

2) $\mathcal{H}_1$ must be invariant under an arbitrary rotation of the reference frame around the z axis. A rotation of the reference frame by an angle ϑ increments the longitudes $\lambda_j, \varpi_j, \Omega_j$ by ϑ. Because $p_j = -\varpi_j$ and $q_j = -\Omega_j$, the invariance of $\mathcal{H}_1$ implies that $\sum_j k_j - \sum_j m_j - \sum_j s_j = 0$. Note that the mean anomaly $l_j = \lambda_j - \varpi_j$ and the argument of perihelion $\omega_j = \varpi_j - \Omega_j$ are invariant by rotation of the reference frame.

3) $\mathcal{H}_1$ must be invariant under a simultaneous change of sign of all inclinations, namely by a transformation $Q_j^{1/2} \to -Q_j^{1/2}, \forall j$ (recall that $Q_j^{1/2} \sim i_j$). This implies that $\sum_j \beta_j/2$ must be an integer number.

4) In the original coordinates and momenta [see (1.29) and (1.38)] $\mathcal{H}_1$ evidently has singularities in real phase space only in correspondence with collisions between two bodies. Therefore, the singularities at $P_j = 0$ and $Q_j = 0$ that $\mathcal{H}_1$ has when written in the form (1.77) are not intrinsic to the function, but must be due to the specific choice of the action–angle variables. The angles p_j and q_j are in fact not defined when the corresponding actions P_j and Q_j are equal to zero, namely when the eccentricity and inclination are null. This singularity, typical of canonical polar coordinates, must be eliminable with the introduction of canonical Cartesian coordinates. We define the *Poincaré variables*

$$y_j = \sqrt{2P_j}\cos p_j \ , \ x_j = \sqrt{2P_j}\sin p_j \ , \ z_j = \sqrt{2Q_j}\cos q_j \ , \ v_j = \sqrt{2Q_j}\sin q_j \ . \tag{1.78}$$

As is easy to check, using the Poisson bracket criterion (1.36), this transformation is canonical and defines x_j and v_j as the new coordinates, and y_j and z_j as the respective conjugate momenta. (y_j, x_j) and (z_j, v_j) are both well defined when $P_j = 0$ and $Q_j = 0$. Therefore in these new canonical variables $\mathcal{H}_1$ must recover its regularity properties. This implies that in (1.77) $\alpha_j - |m_j|$ and $\beta_j - |s_j|$ must be nonnegative even numbers. In fact, only in this case is $\mathcal{H}_1$ an analytic expression of x_j, y_j, z_j, v_j.

1.10 Integrable dynamics

We finally study the dynamics of an integrable system which, as we have seen in Section 1.9, can be generically represented by a Hamiltonian $\mathcal{H}(\mathbf{p})$, independent of the angles $\mathbf{q}$ of the system.

The equations of motion are simply:

$$\dot{p}_j = -\frac{\partial \mathcal{H}}{\partial q_j}(\mathbf{p}) = 0 \ , \quad \dot{q}_j = \frac{\partial \mathcal{H}}{\partial p_j}(\mathbf{p}) \equiv \omega_j(\mathbf{p}) \ , \quad 1 \le j \le n \ , \tag{1.79}$$

from which it follows that the actions p_j are constants of motion, while the angles q_j have constant time derivatives ω_j. Since the coordinates q_j are angles, the time derivatives ω_j are in fact their *frequencies*.

An n-dimensional manifold, which admits as a global coordinate system n independent angles, is called a *torus*, and is usually denoted by $\mathbf{T}^n$. To fix ideas, a one-dimensional torus $\mathbf{T}^1$ is topologically equivalent to a circle, while a two-dimensional torus $\mathbf{T}^2$ is topologically equivalent to the surface of a doughnut. In the motion given by an integrable Hamiltonian, the angles circulate with constant frequencies on tori defined by constant values of the actions $\mathbf{p}$. The tori $\mathbf{p} = \text{constant}$ are therefore *invariant* for the dynamics, in the sense that a trajectory starting on a torus will never leave the torus, the actions $\mathbf{p}$ being constants of motion. The phase space is said to be *foliated in invariant tori*, because every initial condition $\mathbf{p}(0), \mathbf{q}(0)$ generates a motion lying on an invariant torus.

The motion of the angles on a torus depends on the frequencies $\omega_j(\mathbf{p})$. If the frequencies are such that the equation

$$\mathbf{k} \cdot \boldsymbol{\omega} = \sum_{j=1}^{n} k_j \omega_j = 0 \ , \quad \mathbf{k} \equiv (k_1, \ldots, k_n) \in \mathbf{Z}^n \tag{1.80}$$

admits as a unique integer solution $\mathbf{k} = (0, \ldots, 0)$, the motion densely covers the torus. This means that every region of the torus is visited or, more precisely, that, given an arbitrary small neighborhood $U(\mathbf{q}^0)$ of an arbitrary point $\mathbf{q}^0$ on the torus, the motion will sooner or later pass inside $U(\mathbf{q}^0)$. In this case, the frequencies are said to be *nonresonant* and the motion is called *quasi-periodic.*

Conversely, if equation (1.80) admits as solution $n-1$ independent integer nonzero vectors $\mathbf{k}^1, \ldots, \mathbf{k}^{n-1}$, the motion on the torus is *periodic.* In fact, in this case it is easy to see that it is possible to express $n-1$ angles as periodic functions of a unique angle. The frequencies are said to be *completely resonant.*

There are finally intermediate cases where equation (1.80) admits as solution only m independent integer nonzero vectors $\mathbf{k}^1, \ldots, \mathbf{k}^m$, with $m < n-1$. The resulting motion on the torus is neither dense nor periodic. It is possible to express m angles, say for simplicity $q_1, \ldots, q_m$, as periodic functions of the other $n-m$ angles $q_{m+1}, \ldots, q_n$. This implies that the projection of the motion on the $(n-m)$-dimensional torus defined by $q_{m+1}, \ldots, q_n$ is dense, while its projection on the m-dimensional torus defined by $q_1, \ldots, q_m$ is periodic. In this case the frequencies are said to be in a *resonance of multiplicity m.* The quantity

$$\min_{1 \le j \le m} |\mathbf{k}^j| \ , \ \text{with } |\mathbf{k}^j| \equiv |k_1^j| + \cdots + |k_n^j| \tag{1.81}$$

is called the *order of the resonance.*

In general, the frequencies depend on the torus we consider, namely on the values of the actions $\mathbf{p}$. In the case where the Hamiltonian is linear in the actions, however, the frequencies are independent of $\mathbf{p}$. In this case the system is said to be *isochronous*. More generally, a Hamiltonian system is called *degenerate* if the determinant of the matrix of its second-order derivatives vanishes:

$$\det\left(\frac{\partial^2\mathcal{H}}{\partial p_i \partial p_j}\right) = 0 \ . \tag{1.82}$$

If the system is degenerate, then there exists at least one direction in the action space along which the frequencies do not change. Conversely, if the system is nondegenerate, at least one frequency must change for any arbitrary small displacement in the action space. In the latter case, the resonant tori are dense in the phase space, in the sense that arbitrarily close to any point $\mathbf{p}_*$ there is a point $\mathbf{p}$ such that equation (1.80) is satisfied by a nonzero integer vector $\mathbf{k}$.

The theory of Hamiltonian dynamics is well settled for nondegenerate systems and their small perturbations. Conversely, for degenerate systems the properties of the dynamics strongly depend on the specificities of the system under consideration, in particular for what concerns the directions along which the frequencies are preserved, and how the degeneracy is removed when a small perturbation of the system is considered.

Unfortunately, the two-body problem is highly degenerate. As we have seen in the previous section, the Hamiltonian depends only on one action (L) out of three. This implies that the frequencies are the same, whatever the actions G and H are. Moreover, two of the three frequencies, namely $\dot{g}$ and $\dot{h}$ are identically zero, so that g and h are additional constants of motion for the Keplerian problem. Therefore, there is a degeneracy along two directions for $\dot{l}$ and over the entire action space for $\dot{g}$ and $\dot{h}$. For this reason even small perturbations of the two-body problem, like the restricted problem and the planetary problem, may exhibit very complicated dynamics.

Chapter 2

QUASI-INTEGRABLE HAMILTONIAN SYSTEMS

2.1 Introduction to perturbation theory

A Hamiltonian system is said to be *quasi-integrable* if, using a suitable set of canonical action–angle variables, its Hamiltonian function can be written as

$$\mathcal{H}(\mathbf{p},\mathbf{q}) = \mathcal{H}_0(\mathbf{p}) + \varepsilon\mathcal{H}_1(\mathbf{p},\mathbf{q}) \ , \tag{2.1}$$

where ε is a small parameter and $\mathrm{grad}_{\mathbf{p}}\mathcal{H}_0$ and $\mathcal{H}_1$ are intended to be of order unity. It is therefore natural to regard $\mathcal{H}_0$ as the *integrable approximation*, and $\mathcal{H}_1$ as its *perturbation*. In fact, the flow generated by $\mathcal{H}_0$ (namely $\mathbf{p}$ = constant and $\mathbf{q} = \boldsymbol{\omega}_0 t + \mathbf{q}(0)$ with $\boldsymbol{\omega}_0 = \mathrm{grad}_{\mathbf{p}}\mathcal{H}_0$) approximates at order ε the real dynamics generated by $\mathcal{H}$, in the sense that it deviates from the real trajectory by a quantity of order ε in a time of order unity, and by a quantity of order unity in a time of order $1/\varepsilon$. Unfortunately, if one is interested in a very accurate description of the dynamics or in its qualitative behavior on timescales longer than $1/\varepsilon$, knowledge of the flow generated by $\mathcal{H}_0$ is not enough, and one has to look for much better approximations of the real dynamics. This is precisely the goal of *perturbation theory*. As will be detailed in Section 2.5, both the Hamiltonians of the restricted problem (1.70) and of the planetary problem (1.74) have the form (2.1), ε being of the order of the mass of Jupiter relative to that of the Sun, namely $\sim 10^{-3}$. Therefore, the study of quasi-integrable Hamiltonian systems is not just a mathematical challenge, but is of practical importance in Celestial Mechanics.

The power of the Hamiltonian formalism is that, instead of looking for approximations of the real dynamics by working on the equations of motion – which would be very cumbersome – one can work directly on the Hamiltonian

function. More precisely, the general strategy of every perturbation approach for Hamiltonian systems is to look for a canonical transformation *close to the identity* of the form

$$\mathbf{p} = \mathbf{p}^1 + \varepsilon \mathbf{f}_1(\mathbf{p}^1, \mathbf{q}^1) \ , \quad \mathbf{q} = \mathbf{q}^1 + \varepsilon \mathbf{g}_1(\mathbf{p}^1, \mathbf{q}^1) \tag{2.2}$$

such that, by substituting (2.2) in (2.1), the latter becomes

$$\mathcal{H}^1(\mathbf{p}^1, \mathbf{q}^1) = \mathcal{H}_0(\mathbf{p}^1) + \varepsilon \bar{\mathcal{H}}_1(\mathbf{p}^1) + \varepsilon^2 \mathcal{H}_2(\mathbf{p}^1, \mathbf{q}^1) \ , \tag{2.3}$$

with some new functions $\bar{\mathcal{H}}_1$ and $\mathcal{H}_2$ of order unity (typically $\bar{\mathcal{H}}_1$ will be the average of $\mathcal{H}_1$ over the angles $\mathbf{q}$). If this operation is successful, then $\mathcal{H}_0 + \varepsilon \bar{\mathcal{H}}_1$ is the integrable approximation of order ε^2 of the real dynamics. In principle, this procedure can be iterated, looking for a sequence of canonical transformations close to the identity:

$$\mathbf{p}^{r-1} = \mathbf{p}^r + \varepsilon^r \mathbf{f}_r(\mathbf{p}^r, \mathbf{q}^r) \ , \quad \mathbf{q}^{r-1} = \mathbf{q}^r + \varepsilon^r \mathbf{g}_r(\mathbf{p}^r, \mathbf{q}^r) \tag{2.4}$$

such that the Hamiltonian in the action–angle variables $\mathbf{p}^r, \mathbf{q}^r$ becomes

$$\mathcal{H}^r(\mathbf{p}^r, \mathbf{q}^r) = \mathcal{H}_0(\mathbf{p}^r) + \varepsilon \bar{\mathcal{H}}_1(\mathbf{p}^r) + \cdots + \varepsilon^r \bar{\mathcal{H}}_r(\mathbf{p}^r) + \varepsilon^{r+1} \mathcal{H}_{r+1}(\mathbf{p}^r, \mathbf{q}^r) \ , \tag{2.5}$$

thus obtaining a sequence of increasingly better approximations of the real dynamics. One could hope to use this procedure indefinitely, thus transforming the original $\mathcal{H}(\mathbf{p}, \mathbf{q})$ into the integrable $\mathcal{H}^\infty(\mathbf{p}^\infty)$. However, we now know from the work of Poincaré that this hope is vain: in general the procedure cannot be successful up to infinite order (see Section 2.4), but one has to stop at some optimal order r, depending on ε and on the properties of the system. As a consequence, the best integrable approximation of the real dynamics is $\mathbf{p}^r =$ constant, $\mathbf{q}^r = \boldsymbol{\omega}^r t + \mathbf{q}^r(0)$ with $\boldsymbol{\omega}^r = \mathrm{grad}_{\mathbf{p}^r}[\mathcal{H}_0 + \cdots + \varepsilon^r \bar{\mathcal{H}}_r(\mathbf{p}^r)]$. The image of this motion in the original variables $\mathbf{p}, \mathbf{q}$ can be obtained by composing all the sequence of canonical transformations: $\mathbf{p}(t), \mathbf{q}(t)$ have oscillations of size ε around the values $\mathbf{p}^r, \mathbf{q}^r(t)$.

In the following sections we will detail this procedure, and the difficulties that it encounters, following the most frequently adopted approach in Celestial Mechanics: the one making use of Lie series.

2.2 Lie series approach

In the procedure sketched in the previous section, a big problem is to select, among all possible transformations $(\mathbf{p}, \mathbf{q}) \to (\mathbf{p}^1, \mathbf{q}^1)$ that transform (2.1) to (2.3), the one that is canonical. The clever way to avoid the problem (Gröbner,

1960; Hori, 1966; Deprit, 1968) is to look for the good transformation among a class of transformations that are all canonical by construction. This is done by defining the transformations as the time–ε flow of generating Hamiltonians χ, as explained in Section 1.6. More precisely, the transformations are defined as in (1.42), or equivalently as:

$$\mathbf{p} = S_\chi^\varepsilon \mathbf{p}^1 \ , \quad \mathbf{q} = S_\chi^\varepsilon \mathbf{q}^1 \ , \tag{2.6}$$

where S_χ^ε is the Lie series operator, defined in Section 1.7C (see formula (1.52)). All transformations of this kind are indeed canonical. To write the Hamiltonian $\mathcal{H}$ in the new variables, using (1.42) by substitution one gets

$$\mathcal{H}(\mathbf{p}, \mathbf{q}) = \mathcal{H}(\mathbf{p}^1(\varepsilon), \mathbf{q}^1(\varepsilon)) \equiv \mathcal{H}^1(\mathbf{p}^1, \mathbf{q}^1) \ , \tag{2.7}$$

so that $\mathcal{H}^1$ appears as the evolution at time ε of the function $\mathcal{H}(\mathbf{p}^1, \mathbf{q}^1)$ (obtained by formal substitution of $\mathbf{p}, \mathbf{q}$ with $\mathbf{p}^1, \mathbf{q}^1$ in $\mathcal{H}$) along the flow generated by the Hamiltonian χ. As seen in Section 1.7C, $\mathcal{H}^1$ can therefore be written using the Lie series as

$$\mathcal{H}^1 = S_\chi^\varepsilon \mathcal{H} \ . \tag{2.8}$$

The problem is now reduced to finding a suitable generating Hamiltonian χ such that, if $\mathcal{H}$ has the form (2.1), $\mathcal{H}^1$ has the form (2.3). For this purpose, we write (2.8) in explicit form up to order ε^2, using (1.51) and (1.50):

$$\mathcal{H}^1 = \mathcal{H}_0 + \varepsilon \mathcal{H}_1 + \varepsilon\{\mathcal{H}_0, \chi\} + \varepsilon^2\{\mathcal{H}_1, \chi\} + \frac{\varepsilon^2}{2}\{\{\mathcal{H}_0, \chi\}, \chi\} + O(\varepsilon^3) \ . \tag{2.9}$$

In this equation, $\mathcal{H}_0$, $\mathcal{H}^1$ and χ should all be considered as functions of $\mathbf{p}^1, \mathbf{q}^1$. From (2.9), we immediately see that the term of $\mathcal{H}^1$ of order zero in ε is $\mathcal{H}_0(\mathbf{p}^1)$; therefore $\mathcal{H}^1$ will have the form (2.3) if and only if the first-order term in ε will be a function of the sole actions $\mathbf{p}^1$, namely if and only if the equation

$$\mathcal{H}_1 + \{\mathcal{H}_0, \chi\} = \bar{\mathcal{H}}_1 \ , \tag{2.10}$$

which is also called the *homologic equation,* is satisfied by some functions $\chi(\mathbf{p}^1, \mathbf{q}^1)$ and $\bar{\mathcal{H}}_1(\mathbf{p}^1)$.

To solve (2.10) we take advantage of the fact that the coordinates $\mathbf{q}^1$ are angles and that the Hamiltonian $\mathcal{H}$ is periodic in $\mathbf{q}^1$. We expand $\mathcal{H}_1$ in a Fourier series as

$$\mathcal{H}_1(\mathbf{p}^1, \mathbf{q}^1) = \sum_{\mathbf{k} \in \mathbf{Z}^n} c_{\mathbf{k}}(\mathbf{p}^1) \exp\left(\imath \mathbf{k} \cdot \mathbf{q}^1\right) \tag{2.11}$$

where $\mathbf{k}$ is an integer vector, n is the number of degrees of freedom and ι denotes $\sqrt{-1}$. We then look for a solution χ of (2.10) of a similar form

$$\chi(\mathbf{p}^1, \mathbf{q}^1) = \sum_{\mathbf{k}\in\mathbf{Z}^n} d_{\mathbf{k}}(\mathbf{p}^1) \exp\left(\iota\mathbf{k}\cdot\mathbf{q}^1\right) \tag{2.12}$$

so that we can write

$$\{\mathcal{H}_0, \chi\} = -\iota \sum_{\mathbf{k}\in\mathbf{Z}^n} d_{\mathbf{k}}(\mathbf{p}^1)\mathbf{k}\cdot\boldsymbol{\omega}_0(\mathbf{p}^1) \exp\left(\iota\mathbf{k}\cdot\mathbf{q}^1\right) \tag{2.13}$$

where $\boldsymbol{\omega}_0 = \mathrm{grad}_{\mathbf{p}^1}\mathcal{H}_0$. Then, one immediately sees that the solution of equation (2.10) is given by a generating Hamiltonian χ of the form (2.12), with coefficients $d_{\mathbf{k}}$ given by

$$d_{\mathbf{0}} = 0 \ , \quad d_{\mathbf{k}}(\mathbf{p}^1) = -\iota\frac{c_{\mathbf{k}}(\mathbf{p}^1)}{\mathbf{k}\cdot\boldsymbol{\omega}_0(\mathbf{p}^1)} \quad \forall\mathbf{k}\neq\mathbf{0}\ , \tag{2.14}$$

and a function $\bar{\mathcal{H}}_1$ that is simply

$$\bar{\mathcal{H}}_1(\mathbf{p}^1) = c_0(\mathbf{p}^1)\ . \tag{2.15}$$

Now that the generating Hamiltonian χ is determined, the canonical transformation relating the original variables $\mathbf{p}, \mathbf{q}$ to the new ones $\mathbf{p}^1, \mathbf{q}^1$ is also determined. The terms of the new Hamiltonian $\mathcal{H}^1(\mathbf{p}^1, \mathbf{q}^1)$ of quadratic or higher order in ε can be simply computed using the complete expression (2.9).

However, the solution of (2.10) that we have provided is only a *formal* solution. One should check that the generating Hamiltonian χ is well defined as an analytic function, namely that its Fourier series (2.12) with coefficients given by (2.14) is absolutely convergent. In principle, it should also be checked that the power series expansion in ε that defines $\mathcal{H}^1$ in (2.9) is absolutely convergent. However, the latter is guaranteed for ε small enough if χ and $\mathcal{H}$ are analytic. In fact, in this case the right-hand side of (2.9) is by definition the Taylor series expansion of the time evolution of an analytic function ($\mathcal{H}$) along an analytic flow (the one generated by χ), which is known to be convergent provided the time interval $[0, \varepsilon]$ covered by the expansion is short enough.

Therefore, we have only to worry about the analytic properties of χ.

2.3 The small divisors problem

It is immediately evident from (2.12) and (2.14) that the function χ is not defined if there exists an integer vector $\mathbf{k}_*$ such that the denominator

$$\mathbf{k}_*\cdot\boldsymbol{\omega}_0(\mathbf{p}^1) \tag{2.16}$$

vanishes for some $\mathbf{p}^1$ and the corresponding coefficient $c_{\mathbf{k}_*}(\mathbf{p}^1)$ is not zero. Recalling the definition (1.80), this means that the generating Hamiltonian χ is not defined on any torus $\mathbf{p}^1$ that carries *resonant motion* of the integrable approximation $\mathcal{H}_0$. As recalled at the end of Chapter 1, if $\mathcal{H}_0$ is nondegenerate resonances are necessarily dense, so that there is no hope that the function χ can be in general well defined in an open domain of the phase space. Resonances are dense also in most of the interesting degenerate cases, such as the restricted three-body problem. In other words, the perturbation approach outlined in the previous section must fail, in general.

Luckily, there are a couple of ways out of this unpleasant situation. The first way takes advantage of the analytic properties of $\mathcal{H}_1$. A function $f(\boldsymbol{\varphi})$, 2π-periodic in the angles $\boldsymbol{\varphi}$ and analytic on the complex torus

$$\mathbf{T}^n_\sigma = \{\boldsymbol{\varphi} \in \mathbf{C}^n \ : \ \mathrm{Re}\,\varphi_j \in [0, 2\pi] \ , \ |\mathrm{Im}\,\varphi_j| < \sigma \ , \quad 1 \le j \le n\} \tag{2.17}$$

admits a Fourier expansion

$$f(\boldsymbol{\varphi}) = \sum_{\mathbf{k}\in\mathbf{Z}} \alpha_{\mathbf{k}} \exp(\iota \mathbf{k}\cdot\boldsymbol{\varphi}) \ , \quad \text{with } |\alpha_{\mathbf{k}}| < F\exp(-|\mathbf{k}|\sigma) \tag{2.18}$$

where $|\mathbf{k}| = |k_1| + \ldots + |k_n|$ is the *order* of the Fourier harmonic with index $\mathbf{k} = (k_1, \ldots, k_n)$, and F is the supremum of $|f(\boldsymbol{\varphi})|$ for $\boldsymbol{\varphi} \in \mathbf{T}^n_\sigma$ (Arnold, 1963b). Using the exponential decay of its coefficients, it is therefore natural to separate the Fourier expansion of $\mathcal{H}_1$ in two parts, namely $\mathcal{H}_1 = \mathcal{H}_1^{<K} + \mathcal{H}_1^{\ge K}$ with

$$\mathcal{H}_1^{<K} = \sum_{\mathbf{k}\in\mathbf{Z}^n, |\mathbf{k}|<K} c_{\mathbf{k}}(\mathbf{p})\exp(\iota\mathbf{k}\cdot\mathbf{q}) \ , \quad \mathcal{H}_1^{\ge K} = \sum_{\mathbf{k}\in\mathbf{Z}^n, |\mathbf{k}|\ge K} c_{\mathbf{k}}(\mathbf{p})\exp(\iota\mathbf{k}\cdot\mathbf{q}) \ , \tag{2.19}$$

choosing K large enough[1] that $\mathcal{H}_1^{\ge K}$ is of order ε with respect to $\mathcal{H}_1^{<K}$. Then, in the expression (2.9) only $\varepsilon\mathcal{H}_1^{<K}$ appears at order ε, and equation (2.10) admits the solution

$$\chi = \sum_{\mathbf{k}\in\mathbf{Z}^n\setminus\mathbf{0}, |\mathbf{k}|<K} -\iota \frac{c_{\mathbf{k}}(\mathbf{p}^1)}{\mathbf{k}\cdot\boldsymbol{\omega}_0(\mathbf{p}^1)} \exp\left(\iota\mathbf{k}\cdot\mathbf{q}^1\right) . \tag{2.20}$$

Now the series defining χ contains only a finite number of harmonics, so that it will be possible to find an open domain $\mathcal{U}_K$ in the action space, such that

[1]Some authors prefer to choose K independent of ε, thus introducing effectively a new perturbation parameter $\exp(-\sigma K)$. The advantage of this procedure is that the normalized Hamiltonian will be analytic in ε, which is not the case if K is an integer (and therefore discontinuous) function of ε.

the denominators in (2.20) do not vanish for any $\mathbf{p}^1 \in \mathcal{U}_K$. The domain $\mathcal{U}_K$ will be said to be nonresonant up to order K.

A second way to solve the small divisors problem is to choose a point in the action space $\mathbf{p}_*$ such that the frequencies $\boldsymbol{\omega}_0(\mathbf{p}_*)$ satisfy the so-called *diophantine condition*:

$$|\mathbf{k} \cdot \boldsymbol{\omega}_0(\mathbf{p}_*)| > \frac{\gamma}{|\mathbf{k}|^\tau} \ , \ \forall \mathbf{k} \in \mathbf{Z}^n \ , \ \mathbf{k} \neq \mathbf{0} \tag{2.21}$$

with some suitable positive γ and τ. Number theory ensures that these frequencies exist (and the volume of their set in $\mathbf{R}^n$ has positive measure), provided that $\tau > n - 1$. Then we expand $\mathcal{H}_0$ in a power series of $\hat{\mathbf{p}} \equiv \mathbf{p} - \mathbf{p}_*$ as

$$\mathcal{H}_0(\mathbf{p}) = \sum_{j \geq 1} \mathcal{H}_0^j(\hat{\mathbf{p}}) \tag{2.22}$$

with $\mathcal{H}_0^j$ a homogeneous polynomial of degree j in $\hat{\mathbf{p}}$. Finally we consider the action domain of size ε around $\mathbf{p}_*$:

$$\mathcal{U}_\varepsilon(\mathbf{p}_*) = \{\hat{\mathbf{p}} \in \mathbf{R}^n \ : \ \|\hat{\mathbf{p}}\| < \varepsilon\} \tag{2.23}$$

so that the terms $\mathcal{H}_0^j$ with $j \geq 2$ are of higher order in ε with respect to $\mathcal{H}_0^1$. Therefore, in equation (2.9) $\mathcal{H}_0^1$ would play the role of $\mathcal{H}_0$, while the higher-order terms $\mathcal{H}_0^j$ with $j \geq 2$ would be included in $\mathcal{H}_1$. In equation (2.10) the function $\mathcal{H}_0$ in the Poisson bracket with χ would be completely isochronous – being linear in the actions $\hat{\mathbf{p}}$ – with fixed frequencies $\boldsymbol{\omega}_0(\mathbf{p}_*)$, so that the formal solution for χ would be

$$\chi = \sum_{\mathbf{k} \in \mathbf{Z}^n \setminus \mathbf{0}} -\imath \frac{c_{\mathbf{k}}(\mathbf{p}^1)}{\mathbf{k} \cdot \boldsymbol{\omega}_0(\mathbf{p}_*)} \exp\left(\imath \mathbf{k} \cdot \mathbf{q}^1\right) \ . \tag{2.24}$$

Now, using the exponential decay of the coefficients $c_{\mathbf{k}}$ (2.18) and the definition (2.21) of the diophantine frequencies $\boldsymbol{\omega}_0(\mathbf{p}_*)$, one easily sees that (2.24) is absolutely convergent on the complex torus $\mathbf{T}^n_{\sigma-\delta}$, for any positive $\delta < \sigma$.

In practical computations, usually done with the help of an algebraic manipulator implemented on a computer, it is common to use a combination of the two ways discussed above for the solution of the small divisors problem. To be efficient, algebraic manipulators need to represent the functions in power series of the actions and Fourier series of the angles. This naturally leads one to expand $\mathcal{H}_0$ in a power series of $\hat{\mathbf{p}}$, which also has the advantage that the resulting generating Hamiltonian χ (2.24) has denominators that are not functions of the actions, $\mathbf{p}_*$ being fixed. On the other hand, for obvious reasons of limited storage capability, all series expansions need to be truncated. It is natural, then, to keep in $\mathcal{H}_1$ and in χ only the Fourier harmonics of order $|\mathbf{k}| < K$.

2.3.1 Normal forms

The above discussion shows that, although it is not possible in general to transform the original Hamiltonian $\mathcal{H}$ into a new Hamiltonian $\mathcal{H}^1$ of the form (2.3), it is nevertheless possible to achieve this goal if the actions are restricted to a local domain $\mathcal{U}_K$ that is nonresonant up to order K (depending on ε), or to the neighborhood $\mathcal{U}_\varepsilon$ of a torus $\mathbf{p}_*$ with diophantine frequencies. Whenever the Hamiltonian is transformed into the form (2.3) in some domain of the action variables, it is said to be in *Birkhoff normal form* to first order in ε.

However, in some cases, one is interested in a domain that is crossed by resonances of order smaller than K. In these cases, the corresponding terms in the Fourier expansion of $\varepsilon\mathcal{H}_1$ cannot be relegated among the terms of higher order in ε. On the other hand, there is no way to eliminate these terms because of the small divisors problem. Therefore, in these cases it is not possible to transform the Hamiltonian into a Birkhoff normal form, and one has to keep the resonant harmonics in the new Hamiltonian $\mathcal{H}^1$.

More precisely, denote by $\mathcal{U}$ the domain of interest in the action space, and define the *resonant set*

$$\mathcal{K} = \{\mathbf{k} \in \mathbf{Z}^n \ , |\mathbf{k}| < K \ : \ \mathbf{k} \cdot \boldsymbol{\omega}_0(\mathbf{p}) = 0 \text{ for some } \mathbf{p} \in \mathcal{U}\} \ . \tag{2.25}$$

It is then convenient to split $\mathcal{H}_1$ into three parts, namely $\mathcal{H}_1 = \mathcal{H}_1^R + \mathcal{H}_1^{NR} + \mathcal{H}^{\geq K}$ with

$$\mathcal{H}_1^R = \sum_{\mathbf{k}\in\mathcal{K}} c_{\mathbf{k}}(\mathbf{p}) \exp\left(\iota \mathbf{k}\cdot\mathbf{q}\right) \ , \quad \mathcal{H}_1^{NR} = \sum_{\mathbf{k}\in\mathbf{Z}^n, |\mathbf{k}|<K, \mathbf{k}\notin\mathcal{K}} c_{\mathbf{k}}(\mathbf{p}) \exp\left(\iota \mathbf{k}\cdot\mathbf{q}\right) \ , \tag{2.26}$$

and $\mathcal{H}_1^{\geq K}$ defined as in (2.19). With the transformation (2.6), the term of order ε in $\mathcal{H}^1 = S_\chi^\varepsilon \mathcal{H}$ will therefore be $\varepsilon[\mathcal{H}_1^R + \mathcal{H}_1^{NR} + \{\mathcal{H}_0, \chi\}]$. It is then possible to choose the generating Hamiltonian χ such as $\mathcal{H}_1^{NR} + \{\mathcal{H}_0, \chi\} = 0$, setting

$$\chi = \sum_{\mathbf{k}\in\mathbf{Z}^n, |\mathbf{k}|<K, \mathbf{k}\notin\mathcal{K}} -\iota \frac{c_{\mathbf{k}}(\mathbf{p}^1)}{\mathbf{k}\cdot\boldsymbol{\omega}_0(\mathbf{p}^1)} \exp\left(\iota \mathbf{k}\cdot\mathbf{q}^1\right) \ . \tag{2.27}$$

The generating Hamiltonian χ is analytic in $\mathbf{p}^1 \in \mathcal{U}$ because none of the denominators $\mathbf{k}\cdot\boldsymbol{\omega}_0(\mathbf{p}^1)$ vanish for $\mathbf{p}^1 \in \mathcal{U}$ and $\mathbf{k} \notin \mathcal{K}$, by construction of $\mathcal{K}$ (note that the term with $\mathbf{k} = \mathbf{0}$ has also been included in $\mathcal{H}_1^R$), and because the Fourier expansion in (2.27) contains only a finite number of terms.[2] The

[2]Because $\left\|\mathbf{p} - \mathbf{p}^1\right\| = O(\varepsilon)$ (see formula 2.6) to guarantee that $\mathbf{p}^1 \in \mathcal{U}$ one has to choose $\mathbf{p} \in \mathcal{U} - O(\varepsilon)$, where the latter denotes the largest set that is contained in $\mathcal{U}$ together with a neighborhood of radius $O(\varepsilon)$. This implies that the size of $\mathcal{U}$ must be at least $O(\varepsilon)$.

canonical transformation (2.6) is therefore well defined, and the Hamiltonian (2.1) is transformed into

$$\mathcal{H}^1(\mathbf{p}^1,\mathbf{q}^1) = \mathcal{H}_0(\mathbf{p}^1) + \varepsilon\mathcal{H}_1^R(\mathbf{p}^1,\mathbf{q}^1) + O(\varepsilon^2) \ . \tag{2.28}$$

The Hamiltonian $\mathcal{H}^1$ is said to be in *resonant normal form* to first order in ε.

Resonant normal forms are very useful to study resonant dynamics, because they retain at low order only the essential resonant terms, while the nonresonant ones are relegated to higher order. Unfortunately the Hamiltonian $\mathcal{H}_0(\mathbf{p}^1) + \varepsilon\mathcal{H}_1^R(\mathbf{p}^1,\mathbf{q}^1)$ is in general nonintegrable, whenever $\mathcal{K}$ contains more than one resonance,[3] or one resonance with multiplicity larger than 1 (see Section 1.10 for the definition of *multiplicity*). Conversely, in the case where $\mathcal{K}$ contains only one resonance of multiplicity 1, $\mathcal{H}_0(\mathbf{p}^1) + \varepsilon\mathcal{H}_1^R(\mathbf{p}^1,\mathbf{q}^1)$ is integrable, as will be shown in Chapter 4.

2.4 Beyond the first order

The procedure discussed in the previous section allows the elimination of the harmonics with coefficients of order ε that do not belong to the resonant set $\mathcal{K}$. It is then natural to iterate the procedure, in order to eliminate also the nonresonant harmonics with coefficients of higher order in ε. More precisely, one looks for a sequence of generating Hamiltonians χ_r and a sequence of canonical transformations

$$\mathbf{p}^{r-1} = S_{\chi_r}^{\varepsilon^r}\mathbf{p}^r \ , \quad \mathbf{q}^{r-1} = S_{\chi_r}^{\varepsilon^r}\mathbf{q}^r \ , \ r > 1 \tag{2.29}$$

such that finally the Hamiltonian is transformed to the form

$$\mathcal{H}^r(\mathbf{p}^r,\mathbf{q}^r) = \mathcal{H}_0(\mathbf{p}^r) + \varepsilon\mathcal{H}_1^R(\mathbf{p}^r,\mathbf{q}^r) + \cdots + \varepsilon^r\mathcal{H}_r^R(\mathbf{p}^r,\mathbf{q}^r) + \varepsilon^{r+1}\mathcal{H}_{r+1}(\mathbf{p}^r,\mathbf{q}^r) \ , \tag{2.30}$$

with functions $\mathcal{H}_1^R,\ldots,\mathcal{H}_r^R$ containing only harmonics $\mathbf{k}\cdot\mathbf{q}^r$ with $\mathbf{k}\in\mathcal{K}_r$. The latter is the set of resonances up to order K_r that cross the domain $\mathcal{U}$, namely

$$\mathcal{K}_r = \{\mathbf{k}\in\mathbf{Z}^n \ , |\mathbf{k}| < K_r \ : \ \mathbf{k}\cdot\boldsymbol{\omega}_0(\mathbf{p}) = 0 \text{ for some } \mathbf{p}\in\mathcal{U}\} \ , \tag{2.31}$$

where K_r is chosen such that all harmonics with order larger than K_r have coefficients smaller than ε^r (see below). The Hamiltonian (2.30) will be said to be in resonant normal form to order ε^r with respect to $\mathcal{K}_r$. Note that in case $\mathcal{K}_r = \{\mathbf{0}\}$, i.e. no resonances up to order K_r cross the domain of interest,

[3]Meaning that there are at least two values $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathcal{U}$ and two nonzero integer vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ such that $\mathbf{k}_1\cdot\boldsymbol{\omega}_0(\mathbf{p}_1) = \mathbf{k}_2\cdot\boldsymbol{\omega}_0(\mathbf{p}_2) = 0$ and $\mathbf{k}_1\cdot\boldsymbol{\omega}_0(\mathbf{p}_2) \neq 0$, $\mathbf{k}_2\cdot\boldsymbol{\omega}_0(\mathbf{p}_1) \neq 0$.

(2.30) reduces to the form (2.5), and $\mathcal{H}^r$ will be said to be in Birkhoff normal form to order ε^r.

In order to determine the generating function χ_r, one decomposes the perturbation $\mathcal{H}_r$ as $\mathcal{H}_r^R + \mathcal{H}_r^{NR} + \mathcal{H}_r^{\geq K_r}$, where $\mathcal{H}_r^R$ and $\mathcal{H}_r^{NR}$ are defined as in (2.26) and $\mathcal{H}_r^{\geq K_r}$ as in (2.19), with $\mathcal{K}_r$ and K_r instead of $\mathcal{K}$ and K. Remember that, while $\mathcal{H}_r^R$ and $\mathcal{H}_r^{NR}$ contain harmonics whose coefficients are effectively of order ε^r, $\mathcal{H}_r^{\geq K_r}$ contains, by definition of K_r, only harmonic terms with smaller coefficients. Then, analogously to what is done for the resonant normal form of order 1 in ε (see Section 2.3.1), χ_r is chosen to eliminate the nonresonant part of the perturbation of order ε^r, namely as the solution of the equation

$$\mathcal{H}_r^{NR} + \{\mathcal{H}_0, \chi_r\} = 0 \ . \tag{2.32}$$

The key point in this procedure is the choice of K_r. As anticipated above, K_r must be large enough so that all harmonics of order larger than K_r have coefficients smaller than ε^r. Remember that the harmonics of the original perturbation $\varepsilon\mathcal{H}_1$ of order between 1 and K, where K is defined as in (2.19), have coefficients of order ε; therefore, because of the exponential decay (2.18), the harmonics of $\varepsilon\mathcal{H}_1$ of order between $(r-1)K$ and rK have coefficients of order ε^r. As a consequence one must choose $K_r \geq rK$. This implies that the resonant set $\mathcal{K}_r$, must contain more and more resonant terms as r increases. Since all the resonant terms are kept in the normalized Hamiltonian $\mathcal{H}^r$, it turns out that the latter is not in a much more advantageous form than the original Hamiltonian $\mathcal{H}$. In particular, there is in general no hope of constructing Hamiltonians $\mathcal{H}^r$ in Birkhoff normal form to order ε^r, with arbitrarily large r, namely to transform the original Hamiltonian into an integrable $\mathcal{H}^\infty(\mathbf{p}^\infty)$.

In practical applications, one is interested in a given domain $\mathcal{U}$ of the action space, and wants to transform the original Hamiltonian into the simplest possible normal form on that domain. Therefore one usually chooses few independent integer vectors $\mathbf{k}^1, \ldots, \mathbf{k}^m$ corresponding to the main resonances that are present in the domain, and wants to keep in the resonant normal form only the harmonics $\mathbf{k} \cdot \mathbf{q}^r$ with

$$\mathbf{k} \in \mathcal{M} \equiv \left\{ \mathbf{k} \in \mathbf{Z}^n \ : \ \mathbf{k} = n_1\mathbf{k}^1 + \cdots + n_m\mathbf{k}^m \text{ with } (n_1, \ldots, n_m) \in \mathbf{Z}^m \right\} \ . \tag{2.33}$$

It is easy to see that $\mathcal{M}$ – usually called the *resonant module* – is the integer vector space generated by the basis $\mathbf{k}^1, \ldots, \mathbf{k}^m$ (the most interesting case is that where the basis is reduced to one single vector $\mathbf{k}^1$, because the resulting normal form will be integrable – see Chapter 4). Then, one needs to solve equations (2.32) with increasing r, keeping in $\mathcal{H}_r^R$ only the harmonics $\mathbf{k} \cdot \mathbf{q}^r$ with $\mathbf{k} \in \mathcal{M}$. It is evident that this process must stop at some order r, when

the first harmonic that is resonant inside $\mathcal{U}$ and does not belong to the resonant module $\mathcal{M}$ is encountered in $\mathcal{H}_r^{<K_r}$. At that step, the Hamiltonian has already been transformed to the form

$$\mathcal{H}^{r-1} = \mathcal{H}_0 + \mathcal{H}^{r-1}_{\mathrm{norm}} + \varepsilon^r \mathcal{H}_r + O(\varepsilon^{r+1}) \tag{2.34}$$

with $\mathcal{H}^{r-1}_{\mathrm{norm}} \equiv \varepsilon \mathcal{H}_1^R + \cdots + \varepsilon^{r-1}\mathcal{H}^R_{r-1}$. This should be considered as the *optimal normal form*, in the sense that, given $\mathcal{U}$ and $\mathcal{M}$, it minimizes the size of the nonnormalized *remainder* $\varepsilon^r \mathcal{H}_r + O(\varepsilon^{r+1})$. The order r of the optimal normal form and the size of the remainder are completely determined once the domain $\mathcal{U}$ and the resonant module $\mathcal{M}$ are chosen. In fact, the image of $\mathcal{U}$ in frequency space (i.e. the set of frequencies $\boldsymbol{\omega}(\mathbf{p})$ with $\mathbf{p} \in \mathcal{U}$) and $\mathcal{M}$ determine which is the harmonic of lowest order that is resonant in $\mathcal{U}$ and does not belong to $\mathcal{M}$. Denoting by $k_{\min}$ its order, from (2.18) its coefficient cannot be larger than $\exp(-k_{\min}\sigma)$, with some positive σ related to the analytic properties of the original Hamiltonian. As a consequence, in general this harmonic will be encountered in the normalization process at step $r = -k_{\min}\sigma / \ln \varepsilon$. (For more details see Morbidelli and Giorgilli, 1997).

2.4.1 Example of computation of the optimal order of the normal form

Consider the Hamiltonian (2.1) and a point in the action space $\mathbf{p}_*$ such that the frequencies $\boldsymbol{\omega}_0(\mathbf{p}_*)$ satisfy the diophantine property (2.21). Let us construct the optimal Birkhoff normal form in the domain $\mathcal{U}_\varepsilon(\mathbf{p}_*)$ defined in (2.23). Remember that, as explained in footnote 2 (Section 2.3.1), $\mathcal{U}_\varepsilon(\mathbf{p}_*)$ is the domain of minimal size in which one can attempt to construct the normal form. If the Hamiltonian $\mathcal{H}_0$ is nondegenerate, in $\mathcal{U}_\varepsilon(\mathbf{p}_*)$ the frequencies $\boldsymbol{\omega}_0(\mathbf{p})$ cover a neighborhood of size ε of $\boldsymbol{\omega}_0(\mathbf{p}_*)$. Then, taking into account (2.21), in $\mathcal{U}_\varepsilon(\mathbf{p}_*)$ one has

$$|\mathbf{k} \cdot \boldsymbol{\omega}_0(\mathbf{p})| \geq ||\mathbf{k} \cdot \boldsymbol{\omega}_0(\mathbf{p}_*)| - \varepsilon|\mathbf{k}|| \geq \frac{\gamma}{|\mathbf{k}|^\tau} - |\mathbf{k}|\varepsilon \tag{2.35}$$

so that $\mathbf{k} \cdot \boldsymbol{\omega}_0(\mathbf{p})$ can be zero only if $|\mathbf{k}| \geq (\gamma/\varepsilon)^{1/(\tau+1)}$. Therefore, the resonances of lowest order in $\mathcal{U}_\varepsilon(\mathbf{p}_*)$ have order $o \sim 1/\varepsilon^{1/(\tau+1)}$ and the size of the remainder of the optimal Birkhoff normal form is of order $\exp[-\varepsilon^{-1/(\tau+1)}]$, i.e. exponentially small in ε (Morbidelli and Giorgilli, 1997). We can conclude that the dynamics in the ε-neighborhood of an unperturbed diophantine torus is approximated by integrable dynamics, up to a time exponentially large in $1/\varepsilon$.

It is interesting to remark that in the example considered above, if one expands $\mathcal{H}_0$ in Taylor series of $\hat{\mathbf{p}} \equiv \mathbf{p} - \mathbf{p}_*$ as in (2.22), and relegates to higher

order in ε the terms $\mathcal{H}_0^j$ with $j > 1$ (as done in Section 2.3 to construct the Birkhoff normal form to first order in ε), there appears to be no formal obstruction to the construction of the Birkhoff normal form up to any arbitrary order r. In fact, because the leading term $\mathcal{H}_0^1 = \boldsymbol{\omega}_0(\mathbf{p}_*) \cdot \hat{\mathbf{p}}$ is isochronous with diophantine frequencies $\boldsymbol{\omega}_0(\mathbf{p}_*)$, none of the harmonics in the perturbation seem to be resonant in $\mathcal{U}_\varepsilon(\mathbf{p}_*)$, at every order. However, if one computes nontrivial rigorous estimates of the norm of the perturbation, as in Giorgilli and Galgani (1985), one realizes that in this case the norm of $\mathcal{H}_r$ amazingly increases with r as $r!\varepsilon^{-r\tau/(\tau+1)}$, so that $\varepsilon^r\mathcal{H}_r$ behaves as $r!\varepsilon^{r/(\tau+1)}$. This quantity first decreases with r, then definitely diverges when r tends to infinity. The minimal size is achieved when $r \sim \varepsilon^{-1/(\tau+1)}$ and is of magnitude $\exp[-\varepsilon^{1/(\tau+1)}]$. Therefore, one finds again the previous result on the size of the remainder of the optimal normal form.

This example is intended to show that no formal trick may allow the construction of the normal form up to an arbitrarily large order. The presence of resonances inside the domain of interest is an essential obstruction which cannot be passed around.

2.4.2 Generation of high-order harmonics by the normalization process

In many cases of practical importance the original Hamiltonian $\mathcal{H}$ has only a few harmonic terms of low order. In these cases, one could expect that the high-order resonances that cross the domain of interest do not raise the small divisors problem because the coefficients of the corresponding harmonics are identically equal to zero. This would allow the construction of the normal form up to an arbitrarily large order.

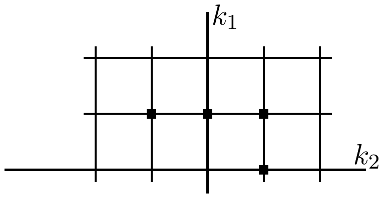
However, this is not true in general. Even if the original Hamiltonian does not have harmonics of high order, these harmonics are usually generated by the construction of the normal form. This can be easily seen in the following example.

Let's consider the two-degree of freedom Hamiltonian (2.1) with

$$\mathcal{H}_0 = \omega_1 p_1 + \omega_2 p_2 + \frac{1}{2}(p_1^2 + p_2^2) \, , \quad \mathcal{H}_1 = \cos q_1 + \cos(q_1 - q_2) + \cos(q_1 + q_2) + \cos q_2 \, , \tag{2.36}$$

with nonresonant ω_1 and ω_2, and try to construct the Birkhoff normal form to first order in ε. The generating Hamiltonian χ is the solution of equation (2.10) and therefore has the same harmonics as $\mathcal{H}_1$. Note that the function $\bar{\mathcal{H}}_1$ in (2.10) is zero, because $\mathcal{H}_1$ has zero average in q_1, q_2.

Let's represent the Fourier spectrum of $\mathcal{H}_1$ and χ as:

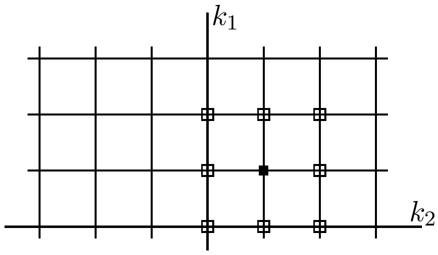


The black dots represent the harmonics $(k_1, k_2) \equiv \exp[\iota(k_1 q_1 + k_2 q_2)]$ which have nonzero coefficient. Because $\mathcal{H}_1$ and χ are real functions, we can restrict our representation to the half-plane $k_1 \geq 0$; in fact, if the coefficient of the harmonic (k_1, k_2) is nonzero, then the coefficient of the harmonic $(-k_1, -k_2)$ must be nonzero, and vice versa.

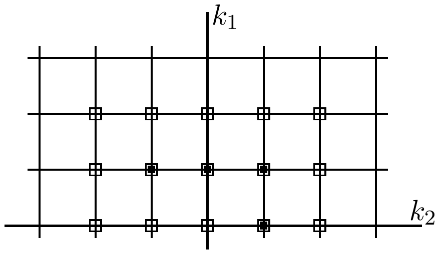
We now proceed to compute which harmonics have nonzero coefficients in the transformed Hamiltonian $\mathcal{H}^1 = S^{\varepsilon}_{\chi}\mathcal{H}$. For this purpose, it is enough to remark how harmonics compose in a Poisson bracket, namely that

$$\begin{aligned}\{\alpha(p_1, p_2)\exp\left[\iota(k_1 q_1 + k_2 q_2)\right], \beta(p_1, p_2)\exp\left[\iota(m_1 q_1 + m_2 q_2)\right]\} \\ = \gamma(p_1, p_2)\exp\left\{\iota[(k_1 + m_1)q_1 + (k_2 + m_2)q_2]\right\}\ ,\end{aligned} \tag{2.37}$$

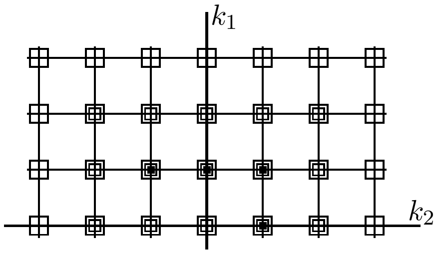
where α and β are generic functions of the actions p_1 and p_2 and the resulting $\gamma = \iota(\alpha\mathbf{k}\cdot\mathrm{grad}_{\mathbf{p}}\beta - \beta\mathbf{m}\cdot\mathrm{grad}_{\mathbf{p}}\alpha)$. For instance, in $\{\mathcal{H}_1, \chi\}$ the harmonic (1,1) in $\mathcal{H}_1$ combined with each harmonic in χ generates with nonzero coefficients all the harmonics that are represented by open squares below



As a consequence, the terms of $\mathcal{H}^1$ of order 2 in ε, namely $\mathcal{L}_{\chi}\mathcal{H}_1$ and $1/2\mathcal{L}^2_{\chi}\mathcal{H}_0$, have the Fourier spectrum represented as in the scheme below:



Similarly, the terms of $\mathcal{H}^1$ of order 3 in ε, namely $1/2\mathcal{L}^2_{\chi}\mathcal{H}_1$ and $1/6\mathcal{L}^3_{\chi}\mathcal{H}_0$, have the Fourier spectrum represented by



This shows that all harmonics are generated with generally nonzero coefficients, despite the fact that the original Hamiltonian had only a few Fourier terms.

It is immediate to see that the generic part of $\mathcal{H}^1$ of size ε^s has harmonics up to order $|\mathbf{k}| = 2s$. Thus the coefficients of the harmonics of order $|\mathbf{k}|$ decay as $\exp(-|\mathbf{k}|\sigma)$, with $\sigma = |\ln\varepsilon|/2$, so that the Fourier law (2.18) is satisfied. Actually this example shows that, after a few normalization steps, the Fourier law is not just an upper bound of the size of the coefficients of the harmonics of the transformed Hamiltonian: the coefficients should be expected to decay no faster than described by the Fourier law. An example with precise computation of the coefficients of the harmonics is reported by Morbidelli and Giorgilli (1997).

2.5 Averaging over the mean motions

The perturbation approach discussed in the previous sections can be directly applied to the Hamiltonians of the restricted problem and of the planetary problem, in the goal of constructing a normal form that is independent of the fast angles (the mean anomalies or, equivalently, the mean longitudes). This procedure is usually (improperly) called *averaging* in the jargon of Celestial Mechanics, because to first order in the planetary masses it is equivalent to a simple average over the fast angles.

For the restricted problem, we start from Hamiltonian (1.70) but, to avoid apparent singularities, we rewrite it using the modified Delaunay variables (1.69). We choose our units of mass, space and time such that $\mathcal{G}m_0 = 1$, and denote by ε the mass of the largest planet, relative to that of the Sun. The vectors $\mathbf{s}_j(t)$ can be written as functions of the planetary orbital elements, and we assume for the moment that the planets move in Keplerian orbits.[4] Thus, the time dependence of the perturbation appears only through the mean longitudes of the planets λ_j. The latter are assumed to move linearly with time, each with its own independent frequency. As explained in Section 1.5, to

[4]It is not necessary at this level to take into account the precession of the planetary orbits, because the precession frequencies are small with ε.

make the Hamiltonian autonomous, it is convenient to overextend the phase space by introducing a couple of conjugate action–angle variables for each independent time–frequency. Therefore, it is natural to choose as new angles the planetary longitudes λ_js, so that, denoting by Λ_j the conjugate actions, the Hamiltonian of the restricted problem has the form (2.1), precisely:

$$\mathcal{H} = \mathcal{H}_0 + \varepsilon\mathcal{H}_1 = -\frac{1}{2\Lambda^2} + \sum_{j=1}^{N} n_j\Lambda_j + \varepsilon\mathcal{H}_1(\Lambda, P, Q, \lambda, p, q, \lambda_1, \ldots, \lambda_N) \quad (2.38)$$

where $n_1, \ldots, n_N$ are the mean motions (i.e. orbital frequencies) of the N planets. The perturbation $\mathcal{H}_1$ is also a function of the planetary longitudes of perihelion p_j and longitudes of node q_j, but these, being assumed fixed, play the role of parameters. The effects produced by their slow motion will be considered in Chapters 8 and 11. Recall from (1.29) that the perturbation $\mathcal{H}_1$ is the sum of the perturbations provided separately by each planet, so that $\mathcal{H}_1 = \sum_j \mathcal{H}_1^{(j)}(\Lambda, P, Q, \lambda, p, q, \lambda_j, p_j, q_j)$. In other words, angles associated to different planets cannot appear in the same harmonic in $\mathcal{H}_1$.

For the planetary problem, we start from Hamiltonian (1.74), rewritten in the modified Delaunay variables (1.76). We denote by ε the mass of the largest planet relative to that of the Sun and by β_j the mass of the j-th planet relative to the mass of the largest one; again the units are chosen so that $\mathcal{G} = m_0 = 1$. With these settings, recalling the definitions of μ_j introduced for (1.39), one has that $\mu_j = \varepsilon\beta_j/(1+\varepsilon\beta_j)$ and $m_0 + m_j = 1 + \varepsilon\beta_j$. Then we define new actions $\Lambda'_j = \Lambda_j/\varepsilon$, $P'_j = P_j/\varepsilon$ and $Q'_j = Q_j/\varepsilon$, where Λ_j, P_j and Q_j are given by (1.76) and keep the angles λ_j, p_j and q_j unchanged. It is easy to see that the Hamiltonian form of the equations of motion is preserved, provided that a new Hamiltonian $\mathcal{H}' = \mathcal{H}/\varepsilon$ is introduced.[5] The new Hamiltonian has the form (2.1), precisely:

$$\begin{aligned}\mathcal{H}' = & \mathcal{H}'_0 + \varepsilon\mathcal{H}'_1 = -\sum_{j=1}^{N} \frac{(1+\varepsilon\beta_j)^{-1}\beta_j^3}{2(\Lambda'_j)^2} \\ & + \varepsilon\mathcal{H}'_1(\Lambda'_1, P'_1, Q'_1, \lambda_1, p_1, q_1, \ldots, \Lambda'_N, P'_N, Q'_N, \lambda_N, p_N, q_N)\ . \quad (2.39)\end{aligned}$$

In the following, for simplicity of notation, we will omit all primes. Recall from (1.38) that $\mathcal{H}_1$ is the sum of functions describing the interaction between a pair of planets. Therefore, in its Fourier expansion, combinations of the angles of more than two planets cannot appear in each harmonic.

To apply the perturbation approach, the perturbations $\varepsilon\mathcal{H}_1$ in (2.38) and (2.39) are expanded in Fourier series and only the harmonics with coefficients

[5] This transformation belongs to an extended class of canonical transformations, not discussed in Section 1.6, where also the Hamiltonian function is transformed.

of order ε are kept in $\varepsilon\mathcal{H}_1$, while those with coefficients of order ε^2 or larger are relegated to a higher-order perturbation function $\varepsilon^2\mathcal{H}_2$, writing $\mathcal{H} = \mathcal{H}_0 + \varepsilon\mathcal{H}_1 + \varepsilon^2\mathcal{H}_2$. The functions $\mathcal{H}_1$ and $\varepsilon\mathcal{H}_2$ play the role of $\mathcal{H}_1^{<K}$ and $\mathcal{H}_1^{\geq K}$ defined in (2.19). The analytic properties of the original Hamiltonian together with the D'Alembert rules ensure that $\mathcal{H}_1$ contains only a finite number of harmonic terms, which can be explicitly computed.

2.5.1 Secular normal form

The normal form of order ε is computed as $\mathcal{H}^1 = S_\chi^\varepsilon \mathcal{H}$, with χ such that equation (2.10) is satisfied. However, the generating Hamiltonian χ cannot be chosen such that the averaged Hamiltonian $\bar{\mathcal{H}}_1$ depends only on the action variables. In fact, because of the degeneracy of the two-body problem, which makes the frequencies of perihelia and nodes identically equal to zero, the harmonics independent of the mean longitudes, but dependent on perihelia and nodes, cannot be eliminated.

To clarify, let's focus on the restricted problem, because the planetary problem is different basically only in the notation. In the restricted problem, because

$$\mathcal{H}_1 = \sum_j \sum_{\substack{k,k_j,m \\ m_j,s,s_j}} c^{(j)}_{k,k_j,m,m_j,s,s_j}(\Lambda, P, Q) \exp\left[\iota(k\lambda + k_j\lambda_j + mp + m_jp_j + sq + s_jq_j)\right] \quad (2.40)$$

the generating Hamiltonian χ would formally be

$$\chi = \sum_j \sum_{\substack{k,k_j,m \\ m_j,s,s_j}} -\iota \frac{c^{(j)}_{k,k_j,m,m_j,s,s_j}(\Lambda^1, P^1, Q^1)}{k\frac{1}{(\Lambda^1)^3} + k_jn_j} \exp\left[\iota(k\lambda^1 + k_j\lambda^1_j + mp^1 + m_jp_j + sq^1 + s_jq_j)\right] , \quad (2.41)$$

where k, k_j, m, m_j, s and s_j are integer numbers, and the new canonical variables $\Lambda = S_\chi^\varepsilon \Lambda^1, \ldots, \lambda = S_\chi^\varepsilon \lambda^1, \ldots$ are those introduced by the Lie series (see formula 2.6). It is then evident that all harmonics with $k = k_j = 0$ would give denominators identically equal to zero in χ. Therefore, these harmonics cannot be eliminated and must be retained in $\bar{\mathcal{H}}_1$. As a consequence, the best normal form that one can hope to construct is of type

$$\mathcal{H}^1 = \mathcal{H}_0(\Lambda^1, \Lambda^1_j) + \varepsilon\bar{\mathcal{H}}_1(\Lambda^1, P^1, Q^1, p^1, q^1, p_j, q_j) + O(\varepsilon^2) \ ; \quad (2.42)$$

the Hamiltonian $\mathcal{H}^1$ will be said to be in *secular normal form* to order ε. In the secular normal form the short periodic terms related to the orbital periods have been averaged out, so that the resulting Hamiltonian describes the long

term or "secular" evolution of the orbit, namely the evolution of P^1 and Q^1 due to the precessional motion of the perihelion p^1 and the node q^1.

The new variables $(\Lambda^1, P^1, Q^1, \lambda^1, p^1, q^1)$ introduced to construct the secular normal form are called the *mean modified Delaunay variables* (of order ε). By construction, they are the action–angle variables for the Hamiltonian of the secular normal form. Formulæ (1.69) allow one to define, from the mean Delaunay variables the *mean orbital elements* $(a^1, e^1, i^1, M^1, \omega^1, \Omega^1)$.

The secular normal form, however, can be constructed only if the denominators $k/(\Lambda^1)^3 + k_j n_j$ that appear in (2.41) are different from zero for each k and k_j appearing in the Fourier expansion of $\mathcal{H}_1$. As discussed in section 2.3, this gives constraints on the domain where the secular normal form can be constructed. More precisely, the expansion of $\mathcal{H}_1$ contains only a finite number of harmonics (k, k_j), the others being relegated in $\varepsilon\mathcal{H}_2$; then the equation $k/\Lambda^3 + k_j n_j = 0$ has a finite number of solutions $\Lambda_0(k, k_j, n_j)$. The secular normal form can be constructed only in the domain, say $\mathcal{U}$, of Λ^1 that excludes suitable neighborhoods of $\Lambda_0(k, k_j, n_j)$. As will be clear in Chapter 4, $\mathcal{U}$ must exclude neighborhoods of size $\sqrt{\varepsilon}$ of Λ_0. If $\Lambda^1 - \Lambda_0 \sim \sqrt{\varepsilon}$, then the generating Hamiltonian χ is also of order $\sqrt{\varepsilon}$, as well as the resulting difference $|\Lambda - \Lambda^1|$ between the original and the new variables. Thus, to ensure that $\Lambda^1 \in \mathcal{U}$, the original actions Λ must be chosen in a domain that is contained in $\mathcal{U}$ together with a neighborhood of radius $|\Lambda - \Lambda^1|$. This situation is sketched in Fig. 2.1. The resonances $k/\Lambda^3 + k_j n_j = 0$ are called *mean motion resonances*, because they concern the mean motions of the body and of the j-th planet. In the planetary problem the mean motion resonances are given by the equation

$$k_i(1 + \varepsilon\beta_i)^{-1}\beta_i^3/\Lambda_i^3 + k_j(1 + \varepsilon\beta_j)^{-1}\beta_j^3/\Lambda_j^3 = 0 \ , \tag{2.43}$$

the indexes i and j referring to two planets.

To construct the secular normal form to higher order in ε one has to iterate the procedure above, determining a sequence of generating Hamiltonians χ_r and a sequence of canonical transformations $((\Lambda^{r-1} = S^{\varepsilon^r}_{\chi_r}\Lambda^r, \ldots, \lambda^{r-1} = S^{\varepsilon^r}_{\chi_r}\lambda^r, \ldots)$ such that the transformed Hamiltonian $\mathcal{H}^r = S^{\varepsilon^r}_{\chi_r}\mathcal{H}^{r-1}$ is in secular normal form up to order ε^r. To this end, χ_r is given by the solution of the equation $\{\mathcal{H}_0, \chi_r\} + \mathcal{H}_r = \overline{\mathcal{H}}_r$, with the function $\overline{\mathcal{H}}_r$ depending only on the actions and on perihelia and nodes. However, as discussed in Section 2.4, $\varepsilon^r\mathcal{H}_r$ must contain all the harmonics whose coefficients are effectively of order ε^r, so that its Fourier expansion contains a number of terms that increases with r. Note moreover that, in the restricted problem, at order ε^2 the term $\{\mathcal{H}_1, \chi\}$ also generates harmonics that depend on the mean longitudes of the body and of *two* planets, while at order ε^3 the term $\{\{\mathcal{H}_1, \chi_1\}, \chi_1\}$ generates harmonics that depend on the mean longitudes of the body and of *three* planets, and

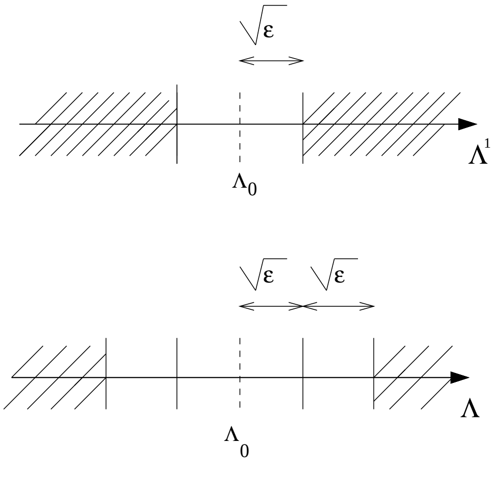


Figure 2.1: Sketch of the geometric construction of the secular normal form. The normal form can be constructed only in the domain with $|\Lambda^1 - \Lambda_0| \gtrsim \sqrt{\varepsilon}$ (the dashed domain in the upper plot). To ensure that Λ^1 falls in such a domain, the original action Λ must be chosen in the domain $|\Lambda - \Lambda_0| \gtrsim 2\sqrt{\varepsilon}$ (the dashed domain in the lower plot); in fact the difference $\Lambda - \Lambda^1$ depends on the angles, and can be as large as $\sim \sqrt{\varepsilon}$.

so forth. Analogously, in the planetary problem the harmonics depending on combinations of the angles of three planets appear at order ε^2, those depending on combinations of the angles of four planets at order ε^3, etc. This leads us to generalize the definition of mean motion resonance to the relations

$$k/\Lambda^3 + \sum_j k_j n_j = 0 \ , \quad \sum_j k_j (1 + \varepsilon\beta_j)^{-1} \beta_j^3 / \Lambda_j^3 = 0 \tag{2.44}$$

for the restricted problem and the planetary problem respectively.

As a consequence of the increasing number of harmonics, the domain on which the secular normal form is constructed must be reduced at each step r, by excluding the neighborhoods of the mean motion resonances (2.44) whose corresponding harmonics appear in $\mathcal{H}_r$. The size of these excluded neighborhoods is $O(\sqrt{\varepsilon})$ in Λ whatever the order r of the normalization step, because $|\Lambda - \Lambda^r| \sim |\Lambda - \Lambda^1| \sim \sqrt{\varepsilon}$ (see Fig. 2.1). Of course, the domain of the original action Λ on which the normal form is constructed must be nonempty. This constrains the total number of resonance neighborhoods that one can exclude. In turn, this forces one to limit the order r at which the secular normal form is constructed.

Unfortunately, because of technical difficulties in the manipulation of the

series, in practical computations the secular normal form is constructed only up to a very limited order in ε. In the restricted problem, in several applications (see for instance Williams, 1969; Williams and Faulkner, 1981; Kozai, 1962; Nakai and Kinoshita, 1985; Yoshikawa, 1987) the secular normal form is computed only to order ε, by simply averaging the perturbation over the mean longitudes. Moreover, since the generating Hamiltonian is not explicitly computed in these works, with some inconsistency the mean elements are simply assumed to be equal to the initial osculating elements.

A more accurate computation has been done by Milani and Knežević (1990), who computed the secular normal form up to order ε^2. To this end, they considered in $\mathcal{H}_1$ the perturbations provided by Jupiter and by Saturn. Moreover, they expanded $\mathcal{H}_1$ in a Fourier series retaining the harmonics with $|k|$ and $|k_j|$ no larger than 14, and expanded its coefficients c_{k,k_j,m,m_j,s,s_j} in powers of the eccentricities and inclinations retaining all terms up to degree 4. This in turn limits, because of D'Alembert rules, the possible values of m, m_j, s and s_j. The mean elements are computed to order ε, using the explicit form of χ, but not to order ε^2, because the second generating Hamiltonian χ_2 is not explicitly computed.

Lemaître and Morbidelli (1994) made a similar computation, but without expanding the perturbation in power series of the eccentricity and inclination of the small body (but still expanding in powers of the eccentricities and inclinations of the planets). For this purpose, on the nodes of a regular grid in (a, e, i) space, they expanded numerically in a Fourier series the perturbation and its derivatives, and constructed the secular normal form to order ε^2. Then, in the applications they computed the secular normal form at any arbitrary point of the (a, e, i) space by interpolation of the results on the grid nodes.

For what concerns the planetary problem, for historical reasons the construction of the secular normal form has not been done following the Hamiltonian approach described in this section. The results are nevertheless equivalent. Laskar (1985, 1986, 1988) computed the equations of motion corresponding to the secular normal form of order ε^2, including all the planets from Mercury to Neptune and expanding the perturbations up to degree 6 in the eccentricities and inclinations. This led him to take into account some 150,000 terms in the series. Bretagnon and collaborators (see for instance Simon and Bretagnon, 1975; Bretagnon 1974, 1982, 1990), in a number of works constructed analytic ephemerides of the planets computing all the orbital elements as Fourier series of time. Their last theory for the four giant planets (still unpublished at the time of writing) is equivalent to computing the secular normal form and the generating Hamiltonians retaining all the terms related to oscillations of eccentricities, inclinations and relative semimajor axes with

amplitude larger than 2×10^{-7}.

2.5.2 Mean motion resonant normal form

Although in practical applications the secular normal form is usually constructed only to a very limited order in ε, still one has to discard the neighborhoods of the main mean motion resonances, namely of the resonances whose corresponding harmonics appear in the perturbation to an order in ε not exceeding the order of the normal form. To study the dynamics in the neighborhood of these resonances, one first has to construct a resonant normal form, following the lines discussed in Section. 2.3.1. For instance, in the restricted problem, in the vicinity of a resonance $k/\Lambda^3 + k_j n_j = 0$, one has to retain in the normal form Hamiltonian not only the secular angles p and q but also of the resonant combination of the mean longitudes $\varphi = k\lambda + k_j\lambda_j$, which would give small denominators in the expression (2.41) of the generating Hamiltonian χ. The latter is therefore determined by solving (2.10), allowing $\bar{\mathcal{H}}_1$ to be a function of p^1, q^1 and φ^1, with $p = S^{\varepsilon}_{\chi} p^1,\ q = S^{\varepsilon}_{\chi} q^1,\ \varphi = S^{\varepsilon}_{\chi} \varphi^1$. More generally, the Hamiltonian is said to be in *mean motion resonant normal form* to order ε^r, if it is transformed to:

$$\mathcal{H}^r_{MMR} = \mathcal{H}_0 + \varepsilon\mathcal{H}^R_1(\Lambda^r, P^r, Q^r, p^r, q^r, \varphi^r) + \ldots \varepsilon^r \mathcal{H}^R_r(\Lambda^r, P^r, Q^r, p^r, q^r, \varphi^r) + O(\varepsilon^{r+1}) \,, \tag{2.45}$$

where $\Lambda^r, P^r, Q^r, p^r, q^r, \varphi^r$ (called the *semimean modified Delaunay variables*) are the new variables related to $\Lambda, P, Q, p, q, \varphi$ through the sequence of Lie series, and φ is the resonant combination of the mean longitudes corresponding to a given resonance. The *semimean orbital elements* are defined from the semimean modified Delaunay variables, by inversion of (1.69).

In practical applications, for mean motion resonances involving the mean longitude of a small body and of a single planet, the mean motion resonance normal form is typically computed only up to first order in ε (see for instance Henrard and Lemaître, 1983, 1987; Sessin and Ferraz-Mello, 1984; Ferraz-Mello, 1987; Yoshikawa, 1990, 1991, for a nonexhaustive list). Conversely for resonances involving the mean longitudes of the small body and of two planets, the normal form is computed up to order ε^2 (Nesvorný and Morbidelli, 1998), because the coefficient of the resonant harmonic is of order ε^2.

In this book we will study the dynamics described by the secular normal form in Chapters 7 and 8 and the dynamics given by the mean motion resonant normal forms in Chapters 9, 10 and 11. But first we need to better understand in general terms the properties of perturbed Hamiltonian dynamics (Chapters 3, 4 and 6) and acquire some useful tools for the numerical exploration of the dynamical properties (Chapter 5).

Chapter 3

KAM TORI

3.1 Kolmogorov's theorem

The considerations of the previous chapter show that it is generically impossible to construct the Birkhoff normal form of a Hamiltonian system up to an arbitrary order in ε on an open domain in the action space, because of the dense presence of resonances. This prevents the global integrability of the system. However, Chapter 2 leaves open the possibility of integrating the motion for specific values of the actions, corresponding to nonresonant frequencies.

A simple Hamiltonian that, although nonintegrable in general, admits one exact solution is

$$\bar{\mathcal{H}}(\mathbf{I},\boldsymbol{\varphi}) = \bar{\mathcal{H}}_0(\mathbf{I}) + \bar{\mathcal{H}}_1(\mathbf{I},\boldsymbol{\varphi}) \quad \text{with } \|\bar{\mathcal{H}}_1\| = O(\|\mathbf{I}\|^2) \ , \tag{3.1}$$

where $\mathbf{I}$ and $\boldsymbol{\varphi}$ are conjugate action–angle variables. In fact, for $\mathbf{I} = \mathbf{0}$, the equations of motion reduce to

$$\dot{\mathbf{I}} = \mathbf{0} \ , \quad \dot{\boldsymbol{\varphi}} = \mathrm{grad}_{\mathbf{I}} \bar{\mathcal{H}}_0(\mathbf{0}) \tag{3.2}$$

from which it is evident that $\mathbf{I} = \mathbf{0}$ and $\boldsymbol{\varphi} = \mathrm{grad}_{\mathbf{I}} \bar{\mathcal{H}}_0(\mathbf{0}) t + \boldsymbol{\varphi}_0$ is the exact solution corresponding to the initial condition $(\mathbf{I},\boldsymbol{\varphi}) = (\mathbf{0},\boldsymbol{\varphi}_0)$. Therefore, the torus $\mathbf{I} = \mathbf{0}, \boldsymbol{\varphi} \in \mathbf{T}^n$ (n being the number of degrees of freedom) is *invariant* for the flow of (3.1), because every orbit initially on the torus will never leave the torus.

In 1954, Kolmogorov proved that, given an analytic quasi-integrable Hamiltonian system $\mathcal{H}(\mathbf{p},\mathbf{q}) = \mathcal{H}_0(\mathbf{p}) + \varepsilon \mathcal{H}_1(\mathbf{p},\mathbf{q})$ and a point $\mathbf{p}_0$ in the action space such that

i) $\boldsymbol{\omega}_0 = \mathrm{grad}_{\mathbf{p}} \mathcal{H}_0(\mathbf{p}_0)$ satisfies the diophantine condition (2.21) with some constants γ and τ;

ii) $\mathcal{H}_0$ is locally nondegenerate in $\mathbf{p}_0$, that is the matrix with entries $(\partial\mathcal{H}_0/\partial p_i \partial p_j(\mathbf{p}_0))$ has nonzero determinant;

there exists a threshold $\bar{\varepsilon}$, depending on $\boldsymbol{\omega}_0$, such that for every $\varepsilon < \bar{\varepsilon}$ a canonical transformation $(\mathbf{p}, \mathbf{q}) \to (\mathbf{I}, \boldsymbol{\varphi})$ allows one to write the Hamiltonian in the form (3.1), with $\mathrm{grad}_{\mathbf{I}}\bar{\mathcal{H}}_0(\mathbf{0}) = \boldsymbol{\omega}_0$.

Kolmogorov's theorem implies that quasi-integrable Hamiltonian systems admit, for small enough perturbations, invariant tori carrying motion with diophantine frequencies. The Hamiltonian equations, although non-integrable in general, can be integrated if restricted to these tori. Kolmogorov's theorem has been later extended and improved by Moser (1962) and Arnold (1963b), and and is therefore now universally known as the *KAM theorem.* The invariant tori carrying nonresonant motion are usually called *KAM tori.*

3.1.1 Sketch of proof of Kolmogorov's theorem

It is interesting and instructive to understand how Kolmogorov's theorem can be proven, since the scheme of the proof is a nice application of the perturbation and normal form theory illustrated in the previous chapter.

The Hamiltonian is first expanded in Taylor series of the actions around $\mathbf{p}_0$, denoting $\hat{\mathbf{p}} = \mathbf{p} - \mathbf{p}_0$. Obviously, $\hat{\mathbf{p}}$ and $\mathbf{q}$ are canonical action–angle variables. In the new variables, the Hamiltonian can be written as

$$\mathcal{H}(\hat{\mathbf{p}}, \mathbf{q}) = \boldsymbol{\omega}_0 \cdot \hat{\mathbf{p}} + \frac{1}{2}\hat{\mathbf{p}} \cdot C\hat{\mathbf{p}} + O(\|\hat{\mathbf{p}}\|^3) + \varepsilon[f_0(\mathbf{q}) + \hat{\mathbf{p}} \cdot \mathbf{f}_1(\mathbf{q}) + O(\|\hat{\mathbf{p}}\|^2)] \quad (3.3)$$

where C is the matrix of the second-order derivatives of $\mathcal{H}_0$ evaluated at $\mathbf{p}_0$, $f_0 \equiv \mathcal{H}_1(\mathbf{p}_0, \mathbf{q})$ and $\mathbf{f}_1$ is a vector with components $\partial\mathcal{H}_1/\partial p_j(\mathbf{p}_0, \mathbf{q})$, $j = 1, \ldots, n$. By $O(\|\hat{\mathbf{p}}\|^j)$ we generically denote terms that are at least of order j in the actions.

We now look for a canonical transformation that gives to (3.3) the form (3.1) up to order ε^2. This is done by composing three canonical transformations, each introduced by a suitable generating Hamiltonian.

The first transformation aims to eliminate the term $f_0(\mathbf{q})$ in the perturbation. We introduce a generating Hamiltonian χ_0 independent of the actions, and the canonical transformation $\hat{\mathbf{p}} = S^{\varepsilon}_{\chi_0}\mathbf{p}_1$, $\mathbf{q} = S^{\varepsilon}_{\chi_0}\mathbf{q}_1$. In the new variables, the Hamiltonian (3.3) becomes

$$\begin{aligned} S^{\varepsilon}_{\chi_0}\mathcal{H} = \boldsymbol{\omega}_0 \cdot \mathbf{p}_1 &+ \frac{1}{2}\mathbf{p}_1 \cdot C\mathbf{p}_1 + O(\|\mathbf{p}_1\|^3) + \varepsilon[\{\boldsymbol{\omega}_0 \cdot \mathbf{p}_1, \chi_0\} + f_0(\mathbf{q}_1) \\ &+ \{\mathbf{p}_1 \cdot C\mathbf{p}_1, \chi_0\} + \mathbf{p}_1 \cdot \mathbf{f}_1(\mathbf{q}_1) + O(\|\mathbf{p}_1\|^2)] + O(\varepsilon^2)\ . \end{aligned} \quad (3.4)$$

It is easy to see that the last two terms in the first row are independent of the actions, while the first two terms in the second row are linear in $\mathbf{p}_1$.

We will denote in the following $\{\mathbf{p}_1 \cdot C\mathbf{p}_1, \chi_0\} + \mathbf{p}_1 \cdot \mathbf{f}_1(\mathbf{q}_1)$ by $\mathbf{p}_1 \cdot \mathbf{f}_{1,1}(\mathbf{q}_1)$. The generating Hamiltonian χ_0 is chosen to be a solution of the equation $\{\boldsymbol{\omega}_0 \cdot \mathbf{p}_1, \chi_0\} + f_0(\mathbf{q}_1) = \bar{f}_0$, with $\bar{f}_0$ denoting the average of f_0 over the angles. The terms independent of the actions annihilate to a constant that can be dropped from the resulting Hamiltonian. Expanding f_0 in a Fourier series as

$$f_0(\mathbf{q}_1) = \sum_{\mathbf{k} \in \mathbf{Z}^n} c_{\mathbf{k}} \exp(\iota \mathbf{k} \cdot \mathbf{q}_1) \ , \tag{3.5}$$

one has

$$\chi_0(\mathbf{q}_1) = -\iota \sum_{\mathbf{k} \in \mathbf{Z}^n \setminus \mathbf{0}} \frac{c_{\mathbf{k}}}{\mathbf{k} \cdot \boldsymbol{\omega}_0} \exp(\iota \mathbf{k} \cdot \mathbf{q}_1) \tag{3.6}$$

where with $\mathbf{k} \in \mathbf{Z}^n \setminus \mathbf{0}$ we denote all nonzero integer vectors. As we have seen in the previous chapter, f_0 being analytic (as ensured by the analytic properties of the original Hamiltonian) and $\boldsymbol{\omega}_0$ diophantine, the Fourier series defining χ_0 in (3.6) converges absolutely.

The second transformation aims to eliminate the new term $\mathbf{p}_1 \cdot \mathbf{f}_{1,1}(\mathbf{q}_1)$ in (3.4). We introduce a generating Hamiltonian χ_1, linear in the actions, and the canonical transformation $\mathbf{p}_1 = S^{\varepsilon}_{\chi_1}\mathbf{p}_2$, $\mathbf{q}_1 = S^{\varepsilon}_{\chi_1}\mathbf{q}_2$. In the new variables, the Hamiltonian (3.4) becomes

$$\begin{aligned} S^{\varepsilon}_{\chi_1} S^{\varepsilon}_{\chi_0} \mathcal{H} = {} & \boldsymbol{\omega}_0 \cdot \mathbf{p}_2 + \frac{1}{2}\mathbf{p}_2 \cdot C\mathbf{p}_2 + O(\|\mathbf{p}_2\|^3) \\ & + \varepsilon[\{\boldsymbol{\omega}_0 \cdot \mathbf{p}_2, \chi_1\} + \mathbf{p}_2 \cdot \mathbf{f}_{1,1}(\mathbf{q}_2) + O(\|\mathbf{p}_2\|^2)] + O(\varepsilon^2) \ . \end{aligned} \tag{3.7}$$

It is easy to see that the first two terms in the second row are both linear in the actions. The generating Hamiltonian χ_1 is determined as a solution of the equation $\{\boldsymbol{\omega}_0 \cdot \mathbf{p}_2, \chi_1\} + \mathbf{p}_2 \cdot \mathbf{f}_{1,1}(\mathbf{q}_2) = \mathbf{p}_2 \cdot \bar{\mathbf{f}}_{1,1}$, the term $\bar{\mathbf{f}}_{1,1}$ denoting the average of $\mathbf{f}_{1,1}$ over the angles. Writing $\chi_1 = \mathbf{p}_2 \cdot \mathbf{g}(\mathbf{q}_2)$ and expanding $\mathbf{f}_{1,1}$ as

$$\mathbf{f}_{1,1}(\mathbf{q}_2) = \sum_{\mathbf{k} \in \mathbf{Z}^n} \mathbf{d}_{\mathbf{k}} \exp(\iota \mathbf{k} \cdot \mathbf{q}_2) \ , \tag{3.8}$$

one has

$$\mathbf{g}(\mathbf{q}_2) = -\iota \sum_{\mathbf{k} \in \mathbf{Z}^n \setminus \mathbf{0}} \frac{\mathbf{d}_{\mathbf{k}}}{\mathbf{k} \cdot \boldsymbol{\omega}_0} \exp(\iota \mathbf{k} \cdot \mathbf{q}_2) \tag{3.9}$$

Again, $\mathbf{f}_{1,1}$ being analytic and $\boldsymbol{\omega}_0$ being diophantine, the series defining $\mathbf{g}$ turn out to be absolutely convergent.

However, this second transformation gives the resulting Hamiltonian:

$$S^{\varepsilon}_{\chi_1} S^{\varepsilon}_{\chi_0} \mathcal{H} = \boldsymbol{\omega}_0 \cdot \mathbf{p}_2 + \frac{1}{2}\mathbf{p}_2 \cdot C\mathbf{p}_2 + O(\|\mathbf{p}_2\|^3) + \varepsilon[\mathbf{p}_2 \cdot \bar{\mathbf{f}}_{1,1} + O(\|\mathbf{p}_2\|^2)] + O(\varepsilon^2) \ , \tag{3.10}$$

where a linear term in the actions $(\mathbf{p}_2 \cdot \bar{\mathbf{f}}_{1,1})$ is still present to order ε. Because of this term, the frequencies of the torus $\mathbf{p}_2 = \mathbf{0}$ (computed neglecting the $O(\varepsilon^2)$ term) are now $\boldsymbol{\omega}_0 + \varepsilon \mathbf{f}_{1,1}$. These frequencies differ from $\boldsymbol{\omega}_0$ only by a quantity of order ε, but this could be enough to make them lose their essential diophantine property, thus forbidding the iteration of the above procedure to eliminate the angular terms that are independent of the actions and linear in the actions at higher order in ε. However, the nondegeneracy of the matrix C in (3.3) – which is guaranteed by condition ii) of the theorem – ensures that close to the torus $\mathbf{p}_2 = \mathbf{0}$ there will be a torus $\mathbf{p}_2 = \varepsilon \boldsymbol{\delta}\mathbf{p}$, with $\|\boldsymbol{\delta}\mathbf{p}\| = O(1)$, with frequencies $\boldsymbol{\omega}_0 + \varepsilon \bar{\mathbf{f}}_{1,1} + \varepsilon C \boldsymbol{\delta}\mathbf{p} = \boldsymbol{\omega}_0$. It is in fact enough to set $\boldsymbol{\delta}\mathbf{p} = -C^{-1}\bar{\mathbf{f}}_{1,1}$, where C^{-1} denotes the inverse of the matrix C (which exists because the determinant of C is different from zero by the nondegeneracy condition). Thus, we introduce a third canonical transformation: $\mathbf{p}_3 = \mathbf{p}_2 - \varepsilon \boldsymbol{\delta}\mathbf{p}$, $\mathbf{q}_3 = \mathbf{q}_2$. This transformation can be written $\mathbf{p}_2 = S^{\varepsilon}_{\chi_{\delta p}} \mathbf{p}_3$, $\mathbf{q}_2 = S^{\varepsilon}_{\chi_{\delta p}} \mathbf{q}_3$, with $\chi_{\delta p} = -\boldsymbol{\delta}\mathbf{p} \cdot \mathbf{q}_3$. As a consequence, in the new action–angle variables the Hamiltonian can be written as

$$S^{\varepsilon}_{\chi_{\delta p}} S^{\varepsilon}_{\chi_1} S^{\varepsilon}_{\chi_0} \mathcal{H} = \boldsymbol{\omega}_0 \cdot \mathbf{p}_3 + \frac{1}{2} \mathbf{p}_3 \cdot C \mathbf{p}_3 + O(\|\mathbf{p}_3\|^3) + \varepsilon O(\|\mathbf{p}_3\|^2) + O(\varepsilon^2) \ . \quad (3.11)$$

Hamiltonian (3.11) has the form (3.1) – the so-called *Kolmogorov normal form* – up to order 2 in ε.

The procedure explained above can be iterated, to give the Hamiltonian the Kolmogorov normal form at increasingly higher order in ε. This is done with a sequence of three transformations of type $\mathbf{p} = S^{\varepsilon^r}_{\chi^{(r)}} \mathbf{p}', \mathbf{q} = S^{\varepsilon^r}_{\chi^{(r)}} \mathbf{q}'$, where $\mathbf{p}, \mathbf{q}, \mathbf{p}', \mathbf{q}'$ generically denote the old and the new action–angle variables and $\chi^{(r)}$ suitable generating Hamiltonians, to eliminate the terms that are independent or linear in the actions at each order r in ε. In this case, there is in principle no obstruction to the iteration of the normalizing procedure up to order infinity. This is because the Kolmogorov normal form does not require the complete elimination of the harmonics – which would be impossible because of the dense presence of resonances – but only the suppression of those whose coefficients are independent of the actions or linear in the actions. The coefficient of each harmonic can still be as large as $O(\varepsilon \|\mathbf{p}\|^2)$. Rigorous estimates show that the series of terms generated by the normalization procedure up to order r converges for $r \to \infty$, provided ε is sufficiently small (see for instance Giorgilli and Locatelli, 1997). This proves Kolmogrov's theorem.

Kolmogorov's theorem is also valid for time-dependent Hamiltonians of type $\mathcal{H}_0(\mathbf{p}) + \varepsilon \mathcal{H}_1(\mathbf{p}, \mathbf{q}, \nu_1 t, \ldots, \nu_n t)$ provided the frequencies of time $\nu_1, \ldots, \nu_n$ are diophantine. As explained in Chapter 1, the Hamiltonian can be written in an autonomous form as in (1.31). Apparently, Kolmogorov's theorem could be not applied to this Hamiltonian, because $\mathcal{H}_0(\mathbf{p}) + \sum_j \nu_j \mathcal{T}_j$ is evidently

degenerate. However, recalling the scheme of the proof, it is evident that, because the actions $\mathcal{T}_j$ do not appear in the perturbation $\mathcal{H}_1$, the frequencies $\nu_1, \ldots, \nu_n$ will never change during the construction of the Kolmogorov normal form. So, it will never be necessary to apply the third transformation – the only one where the nondegeneracy condition plays a role – on the actions $\mathcal{T}_j$. The condition of nondegeneracy of $\mathcal{H}_0(\mathbf{p})$ with respect to the actions $\mathbf{p}$ is therefore sufficient.

3.2 Properties of KAM tori

The KAM theorem gives the first insight into the dynamical properties of quasi-integrable Hamiltonian systems, for small enough perturbations. As illustrated in Chapter 1, the dynamics of the integrable approximation $\mathcal{H}_0(\mathbf{p})$ gives a foliation of the phase space in invariant tori, the actions $\mathbf{p}$ being constants and the angles $\mathbf{q}$ circulating linearly with time, with frequencies $\boldsymbol{\omega}_0(\mathbf{p})$ (left panel in Fig. 3.1). When a small perturbation $\varepsilon\mathcal{H}_1(\mathbf{p}, \mathbf{q})$ is added, the KAM theorem ensures that some tori with diophantine frequencies continue to be invariant for the flow of the complete Hamiltonian $\mathcal{H}_0 + \varepsilon\mathcal{H}_1$. For each invariant torus, new local action–angle variables $\mathbf{I}, \boldsymbol{\varphi}$ can be introduced, such that the Hamiltonian is transformed into the form (3.1). In these variables, the motion on the torus is very simple: the actions $\mathbf{I}$ are constants on the invariant torus and the angles $\boldsymbol{\varphi}$ circulate linearly with time. In the original variables $\mathbf{p}, \mathbf{q}$ the motion on the torus can be computed by composing all the transformations that have been required to give the Hamiltonian the form (3.1). Since each of these transformations is periodic in the angles, the relationship between $\mathbf{p}, \mathbf{q}$ and $\mathbf{I}, \boldsymbol{\varphi}$ is of type $\mathbf{p} = \mathbf{P}(\mathbf{I}, \boldsymbol{\varphi}), \mathbf{q} = \mathbf{Q}(\mathbf{I}, \boldsymbol{\varphi})$, with functions $\mathbf{P}$ and $\mathbf{Q}$ periodic in $\boldsymbol{\varphi}$, and $\mathbf{Q}^{-1}$ periodic in $\mathbf{q}$. Therefore on the torus the angles $\mathbf{q}$ are no longer linear functions of time, but their Fourier spectrum has constant frequencies, while the actions $\mathbf{p}$ have oscillations that are periodic in the angles $\mathbf{q}$, because $\mathbf{p} = \mathbf{P}(\mathbf{I}, \mathbf{Q}^{-1}(\mathbf{q}))$, and quasi-periodic with time, because the angles have nonresonant frequencies. However, the average of $\mathbf{p}$ over time is not equal to the value $\mathbf{p}_0$ for which the unperturbed frequencies $\boldsymbol{\omega}_0(\mathbf{p}_0)$ are equal to the frequencies of the invariant torus. In other words, for a given frequency vector, the perturbed invariant torus is *translated* in phase space with respect to the unperturbed torus. This situation is illustrated in Fig. 3.1. Note that the KAM theorem does not provide any information on the perturbed dynamics away from the invariant tori. This requires the study of resonant dynamics, as will be shown in the next chapter.

The size ε of the perturbation determines which tori continue to be invariant among all the unperturbed ones with diophantine frequencies. Kol-

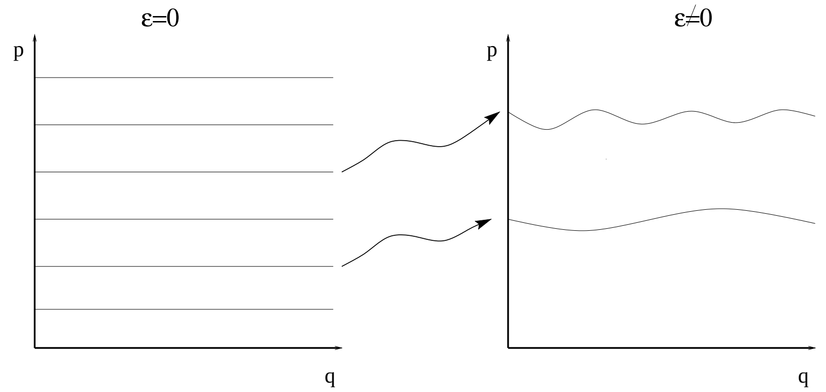


Figure 3.1: For an integrable Hamiltonian the phase space is foliated in invariant tori. According to the KAM theorem, the tori with diophantine frequencies are preserved when a small perturbation is added. However on the perturbed tori the actions are no longer constants of motion, but are periodic functions of the angles. Moreover the mean values of the actions on a KAM torus are in general different from the values of the actions corresponding to the unperturbed torus with the same frequencies.

mogorov's theorem states that ε must be smaller than a threshold $\bar{\varepsilon}$. The latter depends on the frequency vector, through the constants γ and τ that characterize its diophantine property (2.21). In Kolmogorov's theorem, it turns out that $\bar{\varepsilon} \sim \gamma^4$. Therefore, with increasing size of the perturbation, the number of tori that continue to be invariant is reduced, only those with large enough γ being able to survive. If ε is sufficiently large, no invariant tori persist. The KAM theorem, in the version provided by Arnold (1963b), states that the Lebesgue measure of the set of invariant tori is nonzero if ε is sufficiently small. Moreover, this measure grows with decreasing ε, and in the limit $\varepsilon \to 0$ tends to 1. A later estimate by Neishtadt (1982) shows that the measure tends to 1 as $1 - \sqrt{\varepsilon}$. This result can be interpreted from a probabilistic point of view. Let us imagine of making a numerical exploration of a Hamiltonian system and, ignoring everything about its dynamics, to randomly choose the initial conditions of our test orbits. Neglecting round-off errors, the probability of picking up by chance initial conditions corresponding to a KAM torus is proportional to the Lebesgue measure of the set of KAM tori. Therefore, in the light of Arnold's result, in the limit of very small perturbations, the probability of integrating an orbit on a KAM torus is close to 1. The property will be illustrated with numerical examples in the next section.

In Hamiltonian systems with two degrees of freedom, each KAM torus

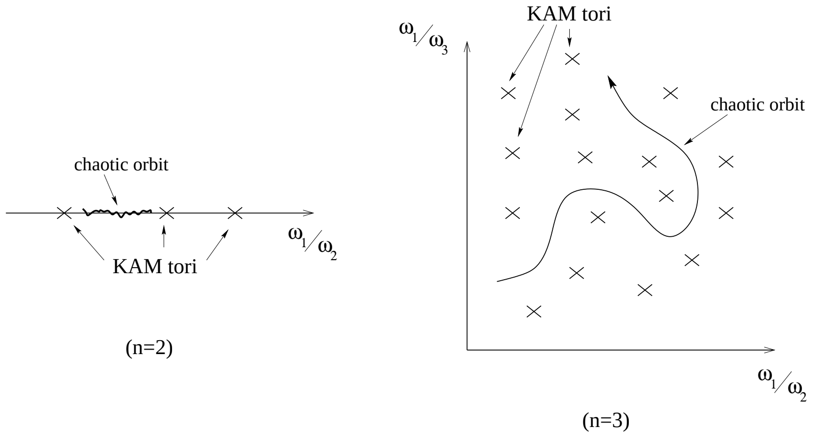


Figure 3.2: An intuitive view of invariant tori and "diffusion" in frequency space: in two degrees of freedom, the frequency space for a given value of the Hamiltonian is a line (left), while in three degrees of freedom it is a plane (right). Invariant KAM tori are invariant points in frequency space. Therefore, in two degrees of freedom, KAM tori bound "diffusion" in frequency space, while in three degrees of freedom chaotic orbits can slalom among tori and "diffuse", in principle, everywhere.

divides the phase space in two parts that are disconnected for the dynamics. In fact, the phase space is four-dimensional, but the conservation of the Hamiltonian forces the motion to evolve in a three-dimensional space. An invariant torus is a two-dimensional manifold, embedded in this two-dimensional space. As a consequence, trajectories cannot pass from one side of a torus to the other without crossing the invariant torus: this is evidently impossible, by definition of invariance (invariant means precisely that the time evolution of any initial condition lying on the torus will never quit the torus). In the next section an illustration of the isolating properties of KAM tori in two degrees of freedom will be given. For Hamiltonian systems with more than two degrees of freedom, the isolating property is no longer true. This is due to the codimension of KAM tori with respect to the phase space. In n degrees of freedom the dimension of the phase space is $2n$. The conservation of the Hamiltonian forces the motion to evolve in a $(2n-1)$-dimensional space. A KAM torus has dimension n; therefore the codimension is $n-1$. This means that the phase space is divided into disconnected parts only if $n = 2$. It is very difficult to "visualize" this important rigorous geometric argument, since one cannot imagine a space and a surface in more than, respectively, three and two dimensions. To have an intuitive nonrigorous view, instead of looking at the dynamics in phase space, it is convenient to restrict attention to

the space of frequencies, which is only $(n-1)$-dimensional for an n-degree of freedom system, once the constancy of the Hamiltonian is taken into account. For a two-degree of freedom system, on a surface of constant Hamiltonian the frequency space is a line, parametrized by the ratios ω_1/ω_2. The KAM tori are fixed points on the frequency line, which correspond to diophantine ratios, and it is evident that each KAM torus cuts the frequency space into two disconnected parts. The trajectories that are not on invariant tori can wander, in principle, on the frequency line, but cannot pass through diophantine ratios (see Fig. 3.2), and are therefore bounded by KAM tori. In a three-degree of freedom system, on a surface of constant Hamiltonian, the frequency space is two-dimensional. We can choose the ratios ω_1/ω_3 and ω_2/ω_3 as coordinates. Again, KAM tori are represented by dots in this frequency plane, while the trajectories that are not on invariant tori may wander on the frequency plane. It is immediately evident from Fig. 3.2 that the latter can slalom among KAM tori and "diffuse", in principle, everywhere.

This portrait, however, is naive: to slalom among KAM tori may take an extremely long time. This can be seen as follows. Once the Hamiltonian is in Kolmogorov normal form (3.1), in the neighborhood of the invariant torus one can construct the Birkhoff normal form, i.e. eliminate all the terms of order r in ε that depend on the angles, with increasing r up to an optimal order. The procedure is completely analogous to that illustrated in Section 2.4.1, so that the optimal size of the remainder of the normal form is determined by the minimal order of the resonances that cross the domain where the normal form is constructed. In the vicinity $\mathcal{U}_\varrho$ of radius ϱ of the invariant torus $\mathbf{I}=\mathbf{0}$ one has:

$$\begin{aligned} |\mathbf{k}\cdot\boldsymbol{\omega}(\mathbf{I})| &= |\mathbf{k}\cdot\boldsymbol{\omega}(\mathbf{0})+\mathbf{k}\cdot(\boldsymbol{\omega}(\mathbf{I})-\boldsymbol{\omega}(\mathbf{0}))| \\ &\geq ||\mathbf{k}\cdot\boldsymbol{\omega}(\mathbf{0})|-|\mathbf{k}||\boldsymbol{\omega}(\mathbf{I})-\boldsymbol{\omega}(\mathbf{0})|| \geq \frac{\gamma}{|\mathbf{k}|^\tau}-|\mathbf{k}|\alpha\varrho\ , \end{aligned} \tag{3.12}$$

where $\boldsymbol{\omega}(\mathbf{I})$ denotes the frequencies of the torus $\mathbf{I}$ in the integrable approximation $\bar{\mathcal{H}}_0(\mathbf{I})$; in the above inequality we have used the diophantine property of the frequencies $\boldsymbol{\omega}(\mathbf{0})$ and the fact that, for each $\mathbf{I}\in\mathcal{U}_\varrho$, $\|\boldsymbol{\omega}(\mathbf{I})-\boldsymbol{\omega}(\mathbf{0})\|\leq\alpha\varrho$, for some positive α. As a consequence, $\mathbf{k}\cdot\boldsymbol{\omega}(\mathbf{I})$ can be zero only if $|\mathbf{k}|\geq(\gamma/\alpha\varrho)^{1/(\tau+1)}$. Therefore, the resonances of lowest order in $\mathcal{U}_\varrho$ have order $o\sim 1/\varrho^{1/(\tau+1)}$, so that the size of the remainder of the optimal Birkhoff normal form will be $\sim\exp[-\varrho^{-1/(\tau+1)}]$.[1] This implies that the "slalom" around

[1]This result can be compared with the one obtained in Section 2.4.1, setting $\varrho=\varepsilon$. The difference is that in Section 2.4.1 the minimal size of the domain where the normal form can be constructed is ε, because the original actions $\mathbf{p}$ have oscillations of order ε. Here, using the actions $\mathbf{I}$, which are constant on the invariant torus, the normal form can be constructed in any neighborhood of arbitrarily small size ϱ.

the invariant torus in Fig. 3.2 must be exponentially slow with the inverse of the distance from the invariant torus. As a matter of fact, Chapter 6 will show that the situation is even more critical: there exists a threshold $\bar{\varrho}$, such that in $\mathcal{U}_{\bar{\varrho}}$ the drift in the frequency space is superexponentially slow. This shows that in more than two degrees of freedom KAM tori, if dense enough, can still provide effective barriers to the "diffusion" of motion in phase space. These barriers can be crossed only after extremely long times.

3.3 Numerical examples

Clear examples of the dynamical structure described by the KAM theorem can be given for two-degree of freedom Hamiltonian systems, because in this case KAM tori can be easily visualized using the so-called *Poincaré sections.*

Poincaré sections are very useful because they allow the representation of the dynamics of a two-degree of freedom system on two-dimensional pictures, one for each value of the corresponding Hamiltonian. This is achieved as follows. Consider the Hamiltonian system $\mathcal{H}(p_1, q_1, p_2, q_2)$, where p_1, q_1 and p_2, q_2 are the canonical action–angle variables, and restrict to the trajectories that satisfy $\mathcal{H}(p_1, q_1, p_2, q_2) = \mathcal{C}$, for some constant $\mathcal{C}$. Choose a two-dimensional surface Σ that is transverse to most such trajectories. Since in general the actions of the system have only small oscillations, while the angles circulate between 0 and 2π, this is usually achieved by choosing Σ as the surface defined by a constant value of one of the angles, for instance setting $q_2 = 0$. For a given value $\mathcal{C}$, the values of p_1, q_1 on the surface $q_2 = 0$ unequivocally determine the value of the remaining action p_2, which can be computed by solving the implicit equation $\mathcal{H}(p_1, q_1, p_2, 0) = \mathcal{C}$. Now, by numerically integrating the equations of motion (see for instance Press *et al.*, 1986), compute the successive intersection that each trajectory has with the surface Σ and consider only those generated when the trajectory pierces the surface in a given direction, for instance with $\dot{q}_2 > 0$. The sequence of points p_1, q_1 that each trajectory leaves on Σ gives an unequivocal stroboscopic image of the time evolution of the trajectory in phase space.

If a trajectory lies on a KAM torus, the sequence of points p_1, q_1 must fall on a one-dimensional curve. In fact, on a KAM torus the actions are periodic functions of the angles, so that – denoting by $p_1 = P(q_1, q_2)$ the periodic relationship between p_1 and both q_1 and q_2 – on the surface Σ defined by $q_2 = 0$ the points p_1, q_1 must lie on the curve $p_1 = P(q_1, 0)$. Moreover, because the angles have nonresonant frequencies (see the definition of nonresonance in Section 1.8), at each intersection with Σ the angle q_1 has to assume a different value on the interval $[0, 2\pi]$ and, with time passing, the sequence of points on

Σ must densely fill the curve $p_1 = P(q_1, 0)$. As a consequence, if the Poincaré section is computed for long enough time, the KAM tori appear as solid curves crossing the entire interval $[0, 2\pi]$.

By way of example, instead of writing a two-degree of freedom Hamiltonian system and computing the Poincaré section, it is much simpler to use the so-called *standard map*, where the sequence of points p_1, q_1 is explicitly given iterating the relationships

$$q_1' = q_1 + p_1 \ , \quad p_1' = p_1 + \varepsilon \sin(q_1') \ . \tag{3.13}$$

The standard map can be regarded as the Poincaré section of a two-degree of freedom Hamiltonian (Henrard, 1970) – although the latter cannot be written in explicit form – so that the considerations above on the representation of KAM tori in Poincaré sections apply also to the standard map. If $\varepsilon = 0$ in (3.13) the standard map reduces to the Poincaré section at $q_2 = 0$ of the integrable Hamiltonian system $\mathcal{H}(p_1, q_1, p_2, q_2) = p_1^2/2 + 2\pi p_2$, so that ε plays the role of a perturbation parameter.

We now show the portraits of the standard map for different values of ε (Fig. 3.3) and discuss them in the light of KAM theory. The reader may easily implement the standard map equations on his own computer and familiarize himself with the dynamical structures related to the different values of the perturbation parameter.

In Fig. 3.3 we have chosen 20 initial conditions on the axis $q_1 = 0$, with initial p_1 regularly spaced on the interval $[-\pi, \pi]$. In the case where ε is small (panel a) every initial condition generates an orbit lying on a KAM torus. This implies that the volume filled by KAM tori is large. Note however that the tori are significantly distorted with respect to the case $\varepsilon = 0$, where p_1 would be constant on each torus. The distortion of the KAM tori leaves an "empty region" around $p_1 = 0$. This is a *resonance region* (associated to the resonance $\dot{q}_1 = 0$), and the trajectories there do not cross the $q_1 = 0$ axis, so that they cannot be computed with our choice of initial conditions. The dynamics in the resonance region will be investigated in detail in the next chapter. Finally note that, while most of the KAM tori appear as solid curves in panel (a), the one passing through $p_1 = -2.24$ on the $q_1 = 0$ axis (and, by symmetry, also the one passing through $p_1 = 2.24$) appears as a dotted curve. This is because the time required to fill the torus in a dense way with respect to the graphic resolution depends on the frequencies on the torus. The closer the frequencies are to a rational ratio, the longer is this time. In Fig. 3.3 only 1000 iterations of the standard map are computed for each initial condition. If the number of iterations were doubled, also this torus would appear as a solid curve.

In panel (b), the value of ε is increased by a factor of 3. The dynamical structure changes significantly. Now only 14 initial conditions generate

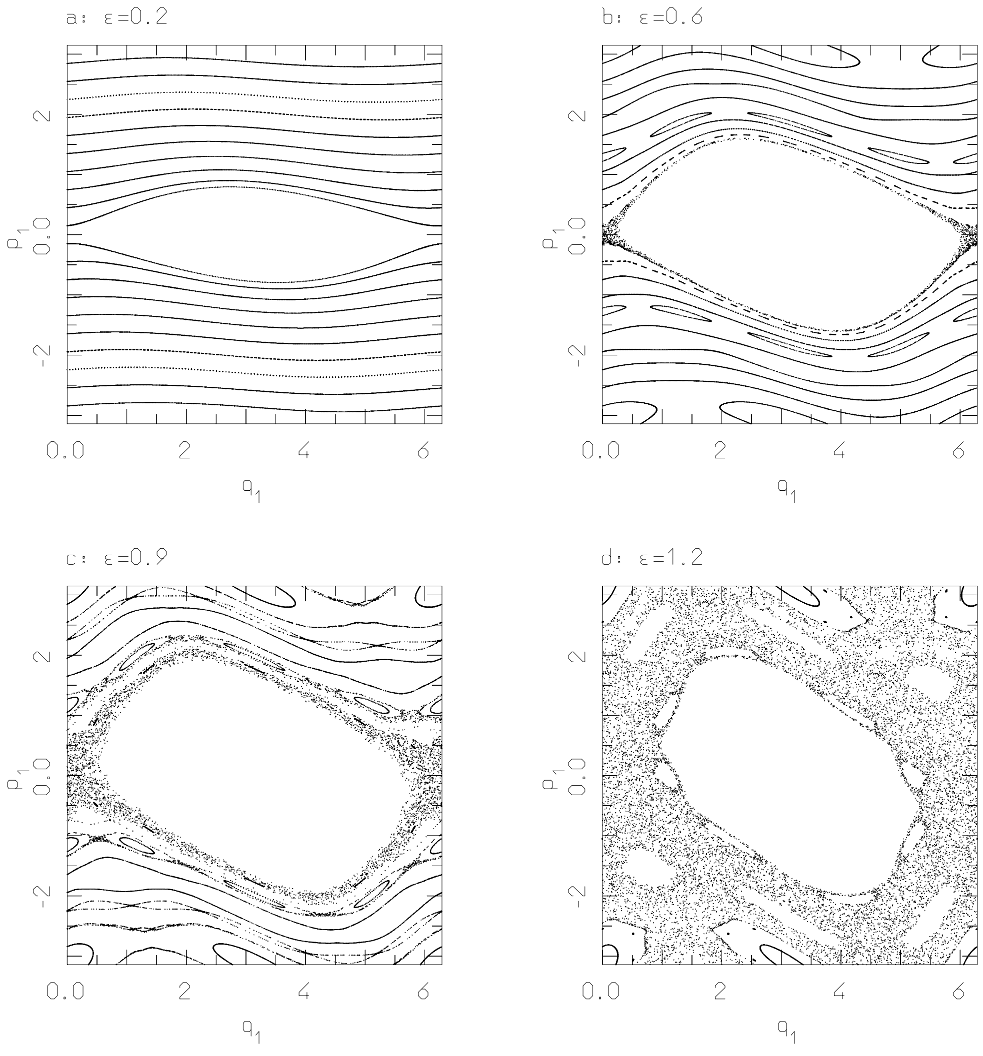


Figure 3.3: The phase portrait of the standard map for different values of ε. Note that with increasing ε the KAM tori become more distorted and more rare, while chaotic regions become more and more extended. See the text for comments.

trajectories lying on KAM tori. The initial conditions with $p_1 = \pm 2.84$ and $p_1 = \pm 1.36$ generate trajectories still lying on invariant curves, but these curves have a topological structure that is different from that of KAM tori: they appear as chains of closed circles (islands). These invariant curves are also related to resonant dynamics and they will be discussed in the next chapter. Finally, the trajectories with initial conditions $p_1 = \pm 0.15, q_1 = 0$ do not lie on invariant curves, but appear as a collection of scattered points. The region covered by the these points is called the *chaotic region*, and we will see that its existence is also related to resonant dynamics. In the chaotic region KAM tori do not exist, and the actions are not periodically related to the angles. Note that the distortion of the invariant tori and the size of the "empty region" at the center of the figure have both increased with respect to panel (a).

In panel (c), only four initial conditions generate trajectories on KAM tori, showing that the volume filled by KAM tori shrinks with increasing value of ε. Chaotic zones appear also around some chains of islands, and not only around the central "empty region" as was the case in panel (b). As we said in the previous section, the surviving KAM tori separate the (p_1, q_1) space into disconnected regions, bounding the dynamical evolution of the trajectories in the chaotic regions. For instance, the trajectory starting with $p_1 = -0.15$ will never reach the region with $p_1 < -2.5$, because a KAM torus should be crossed.

Finally, in panel (d), all KAM tori have disappeared. The chaotic regions have merged together, and now dominate the phase portrait of the system. It is not possible to distinguish the evolutions of different initial conditions in the chaotic zone. The action p_1 can assume any value during the evolution of chaotic trajectories.

From the examples above, it is clear that the KAM theorem is able to give a satisfactory description of the global dynamical structure of the system only for very small values of the perturbation parameter. When the latter increases, KAM tori become more and more rare, and other structures appear. To have a complete understanding of Hamiltonian dynamics it is therefore necessary to investigate the complement of KAM tori, that is resonant dynamics and interactions among resonances. This will be done in the next chapter and in Chapter 6.

Chapter 4

SINGLE RESONANCE DYNAMICS

4.1 The integrable approximation

We now consider the n-degree of freedom quasi-integrable Hamiltonian $\mathcal{H}(\mathbf{p},\mathbf{q}) = \mathcal{H}_0(\mathbf{p}) + \varepsilon\mathcal{H}_1(\mathbf{p},\mathbf{q})$ and concentrate our attention on a domain $\mathcal{U}$ in the action space characterized by *one* main resonance of multiplicity 1 (see Section 1.10 for the definition of multiplicity). Our study of resonant dynamics starts by constructing the resonant normal form, in order to reduce to a minimal size all the harmonics that are not related to the main resonance. More precisely, we denote by $\bar{\mathbf{k}}$ the minimal nonzero integer vector related to the main resonance, namely the vector of minimal norm such that $\bar{\mathbf{k}}\cdot\omega_0(\mathbf{p}) = 0$ for some $\mathbf{p} \in \mathcal{U}$, and we construct the resonant normal form with respect to the resonant module $\mathcal{M}$ generated by $\bar{\mathbf{k}}$ (see definition 2.33).

Following Chapter 2, the resonant normal form of the Hamiltonian is of type:

$$\mathcal{H} = \mathcal{H}_0(\mathbf{p}) + \mathcal{H}_{\rm norm}(\mathbf{p}, \bar{\mathbf{k}}\cdot\mathbf{q}) + \mathcal{R}(\mathbf{p},\mathbf{q}) \ , \tag{4.1}$$

where for simplicity we still denote by $\mathbf{p},\mathbf{q}$ the new action–angle variables introduced for the construction of the normal form. The normalized perturbation $\mathcal{H}_{\rm norm}$, whose size is not larger than ε, contains only the main resonant harmonic $\bar{\mathbf{k}}\cdot\mathbf{q}$ and its integer multiples; $\mathcal{R}$ denotes the remainder of the optimal normal form, whose size will be discussed in Section 4.3.

For the moment, we neglect the presence of the remainder $\mathcal{R}$ and consider only the truncated resonant Hamiltonian $\mathcal{H}_{\rm res} = \mathcal{H}_0 + \mathcal{H}_{\rm norm}$. To study the dynamics of $\mathcal{H}_{\rm res}$ it is convenient to introduce a new angle $\varphi_1 = \bar{\mathbf{k}}\cdot\mathbf{q}$. This can be done with a canonical transformation. In fact, it can be shown (see Lemma 4 in Morbidelli and Giorgilli, 1993) that it is always possible to define

a matrix U which has integer entries, determinant equal to 1, and first row equal to $\bar{k}_1, \ldots, \bar{k}_n$ (components of $\bar{\mathbf{k}}$); then, denoting by $(U^T)^{-1}$ the inverse of the transposition of U, the transformation defined by $\boldsymbol{\varphi} = U\mathbf{q}$, $\mathbf{I} = (U^T)^{-1}\mathbf{p}$ is canonical, and it gives by construction $\varphi_1 = \bar{\mathbf{k}} \cdot \mathbf{q}$. The angle φ_1 is usually called the *resonant angle* or *critical angle of the resonance.* In the new variables, $\mathcal{H}_{\rm res}$ has the form:

$$\mathcal{H}_{\rm res}(\mathbf{I}, \boldsymbol{\varphi}) = \mathcal{H}_0(\mathbf{I}) + \mathcal{H}_{\rm norm}(\mathbf{I}, \varphi_1) \ . \tag{4.2}$$

Because the system depends only on one angle (φ_1), the Hamiltonian $\mathcal{H}_{\rm res}$ is trivially integrable.[1] The actions $I_2, \ldots, I_n$ are constants of motion, so that the interesting part of the dynamics takes place on the (I_1, φ_1) plane. On this plane, the motion evolves along level curves of $\mathcal{H}_{\rm res}(I_1, \varphi_1)$, the other actions playing the role of parameters.

Generically, on the (I_1, φ_1) plane the dynamics of $\mathcal{H}_{\rm res}$ has the *local* structure of the dynamics of a pendulum. This can be seen as follows. Denote by $\mathbf{p}^*$ a value of the actions $\mathbf{p}$ that is exactly resonant, namely satisfying the equation $\bar{\mathbf{k}} \cdot \boldsymbol{\omega}_0(\mathbf{p}^*) = 0$. Let $\mathbf{I}^* = (U^T)^{-1}\mathbf{p}^*$. Fix now in (4.2) $I_2, \ldots, I_n$ equal to $I_2^*, \ldots, I_n^*$, and expand the Hamiltonian in a Taylor series of I_1 around I_1^*. Defining $\hat{I}_1 = I_1 - I_1^*$ and neglecting terms of order $\hat{I}_1^3$ in $\mathcal{H}_0$ and of order $\hat{I}_1$ in $\mathcal{H}_{\rm norm}$, (4.2) becomes:

$$\mathcal{H}_{\rm res} = \alpha(\mathbf{I}^*)\hat{I}_1 + \frac{\beta(\mathbf{I}^*)}{2}\hat{I}_1^2 + \mathcal{H}_{\rm norm}(\mathbf{I}^*, \varphi_1) \ , \tag{4.3}$$

where α and β are the coefficients of the two first orders of the Taylor expansion of $\mathcal{H}_0$. In the expression above, α must necessarily be equal to zero; in fact, taking into account only $\mathcal{H}_0$, α is equal to $\dot{\varphi}_1$ at the resonant location $\hat{I}_1 = 0$, but $\dot{\varphi}_1 = \bar{\mathbf{k}} \cdot \dot{\mathbf{q}} = \bar{\mathbf{k}} \cdot \boldsymbol{\omega}_0(\mathbf{p}^*) = 0$. Finally in (4.3) expand $\mathcal{H}_{\rm norm}$ in a Fourier series of φ_1 and retain only the main term. Assuming, without loss of generality, that the leading harmonic is $c(\mathbf{I}^*) \cos\varphi_1$, the Hamiltonian $\mathcal{H}_{\rm res}$ takes the form

$$\mathcal{H}_{\rm res} = \frac{\beta}{2}\hat{I}_1^2 + c\cos\varphi_1 \ , \tag{4.4}$$

which is the well-known Hamiltonian of the pendulum.

The left panel of Fig. 4.1 shows the dynamics of the pendulum, assuming β and c both positive; the left and the right border of the panel should be identified. The pendulum has two equilibrium points both at $\hat{I}_1 = 0$: one at $\varphi_1 = 0$ – unstable or *hyperbolic* – the other at $\varphi_1 = \pi$ – stable or *elliptic* –.

[1] If the considered resonance had multiplicity $m > 1$, the resulting truncated normal form would not be integrable, because in this case it would depend on m independent angles. The case of resonances of larger multiplicity will be discussed in Chapter 6.

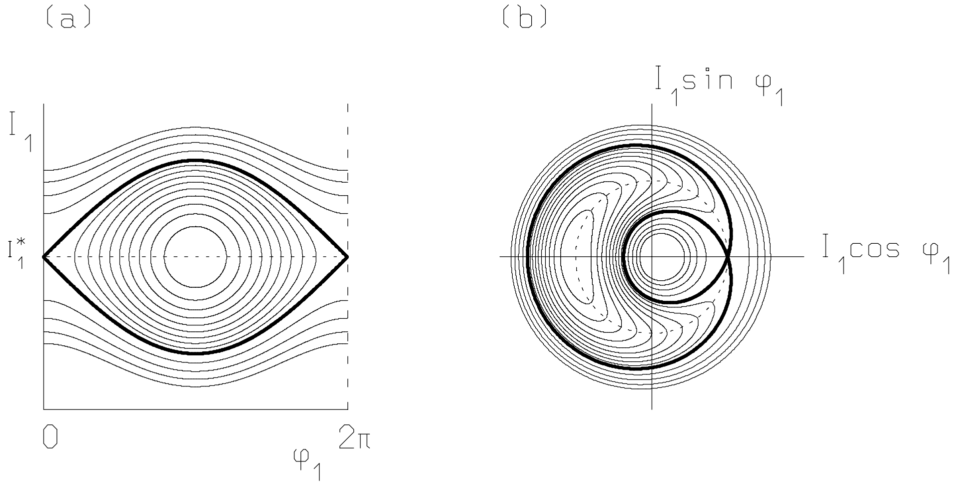


Figure 4.1: Integrable resonant dynamics in local Cartesian coordinates (left panel) and global polar coordinates (right panel). The bold curve denotes the separatrix; the dotted line/circle denotes the unperturbed resonant value $I_1 = I_1^*$.

The bold curve, connecting the unstable equilibrium point to itself (module 2π), is called the *separatrix*, because it separates the $(\hat{I}_1, \varphi_1)$ plane in three regions with different dynamical properties. Above the separatrix, the angle φ_1 circulates from 0 to 2π with positive derivative; its frequency monotonically increases with the distance from the separatrix. Below the separatrix, φ_1 circulates again, but with negative derivative. Inside the separatrices, φ_1 *librates* around π, and the motion evolves on a closed curve surrounding the stable equilibrium point. The frequency of libration is $\sqrt{c\beta}$ for the trajectory closest to the stable equilibrium, and decreases to zero approaching the separatrix. On the separatrix it takes an infinite time to travel from the unstable equilibrium to itself, because the motion slows down exponentially with time, approaching the unstable equilibrium. The regions above and below the separatrix are called the *circulation regions*, while the region inside the separatrix is called the *libration region.* The resonant region is, strictly speaking, the libration region, because only for librating orbits is the averaged time derivative of φ_1 equal to zero. Conversely, above and below the separatrix $\dot{\varphi}_1$ is on average positive or negative. The half-width of the resonant region, measured as the value of $\hat{I}_1$ at the apex of the separatrix, is $2\sqrt{c/\beta}$. The minimal size of the domain $\mathcal{U}$ for the construction of the resonant normal form, therefore, must be $4\sqrt{c/\beta}$ in I_1. We will see in Section 4.3 how this determines the size of the remainder of the optimal normal form. The half-width of the resonance

in the unperturbed frequency space is equal to the unperturbed frequency of the resonant angle φ_1 at the apex of the separatrix, namely $2\sqrt{\beta c}$.

Of course, (4.4) is only the *local* approximation of the dynamics of $\mathcal{H}_{\rm res}$ around $I_1 = I_1^*$. But in most problems of Celestial Mechanics I_1 and φ_1 constitute a system of *global polar coordinates.* The *global* representation of the resonant dynamics in these coordinates is shown in the right panel of Fig. 4.1. As one sees, around $I_1 = I_1^*$ the dynamics is topologically equivalent to that shown on the left panel; however close to the coordinate center $I_1 = 0$ a third (stable) equilibrium point appears. In fact, introducing the canonical Cartesian coordinates $x = \sqrt{2I_1}\sin\varphi_1$, $y = \sqrt{2I_1}\cos\varphi_1$, if the Hamiltonian $\mathcal{H}_{\rm res}$ is analytic at $x = y = 0$ (as ensured by the D'Alembert rules for the Hamiltonians of Celestial Mechanics), all curves must be smooth; therefore, close to the center the curves must be topologically equivalent to circles, surrounding a fixed point. The analytic property in $I_1 = 0$ implies that the Fourier expansion of (4.2) has the form $\sum_m c_m(I_1)\exp[\imath m\varphi_1]$, with $c_m(I_1) \sim I_1^{m/2+k}$, where k is a nonnegative integer number (see the discussion for rule No. 4 in Section 1.9.3). If $c_1(I_1) = \sqrt{I_1}$ (i.e. $k = 0$), then the equilibrium point is offset with respect to $I_1 = 0$, while in all other cases it is at $I_1 = 0$.

The portrait of Fig. 4.1 can be considered as a paradigm for integrable resonance dynamics. In fact, higher-order harmonics in φ_1 and the complete dependence on I_1 of all coefficients can deform the dynamical portrait of $\mathcal{H}_{\rm res}$ with respect to the picture shown in the right panel of Fig. 4.1 but do not change, in general, its basic features: the presence of a stable equilibrium point surrounded by a separatrix that connects an unstable equilibrium point to itself, and a second stable equilibrium close to the origin of the polar coordinates.

4.2 Resonant action–angle variables

Because the truncated resonant normal form (4.2) is integrable, following the Arnold–Liouville theorem one can introduce new action–angle variables $\mathbf{J}, \boldsymbol{\psi}$ and write $\mathcal{H}_{\rm res}$ solely as a function of the actions $\mathbf{J}$. We will denote hereafter these new variables as *resonant action–angle variables.* The introduction of resonant action–angle variables is of crucial importance in Celestial Mechanics to study in detail the perturbed resonant dynamics. Therefore, in this section we detail how these variables are defined and how the new Hamiltonian can be handled using numerical techniques.

Since in (4.2) the actions $I_2, ..., I_n$ are constants of motion, the new actions $J_2, ..., J_n$ are simply equal to the original ones: therefore $J_2 = I_2, ..., J_n = I_n$. The definition of J_1 is more complicated. Having fixed the values of $I_2, ..., I_n$,

the Hamiltonian (4.2) is reduced to one degree of freedom. So, in I_1, φ_1, every trajectory is represented by a closed curve. These curves are the cycles that are used in Arnold's recipe for the definition of the new action. As Fig. 4.1 shows, in a resonant problem there are circulation cycles (where φ_1 ranges from 0 to 2π) and libration cycles (where φ_1 is bounded between $\varphi_{1\mathrm{min}}$ and $\varphi_{1\mathrm{max}}$). On each cycle, one can express the action I_1 as a function of φ_1, by solving the implicit equation $\mathcal{H}_{\mathrm{res}}(I_1, \varphi_1) = E$, the constant E characterizing the value of the Hamiltonian on the cycle. As explained in Section 1.9, the action J_1 is defined by the integral over the cycle of the action I_1, namely:

$$J_1 = \frac{1}{2\pi} \oint I_1(E, \varphi_1) \mathrm{d}\varphi_1 \ . \tag{4.5}$$

For a circulation cycle, this is simply

$$J_1 = \frac{1}{2\pi} \int_0^{2\pi} I_1(E, \varphi_1) \mathrm{d}\varphi_1 \tag{4.6}$$

which is (normalized by 2π) the area enclosed between the cycle and the $I_1 = 0$ axis (see Fig. 4.2, panel a). For a libration cycle, (4.5) reads

$$J_1 = \frac{1}{2\pi} \left[\int_{\varphi_{1\mathrm{min}}}^{\varphi_{1\mathrm{max}}} I_1^+(E, \varphi_1) \mathrm{d}\varphi_1 - \int_{\varphi_{1\mathrm{min}}}^{\varphi_{1\mathrm{max}}} I_1^-(E, \varphi_1) \mathrm{d}\varphi_1 \right] \ , \tag{4.7}$$

where I_1^+ denotes the upper part of the cycle and I_1^- the lower part. The difference between the two integrals gives the dotted area shown in Fig. 4.2 (panel b), i.e. the area enclosed by the libration cycle.

For what concerns the new angles $\boldsymbol{\psi}$, these can be introduced via the generating function $S = \int \sum_{i=1}^n I_i(\mathbf{J}, \boldsymbol{\varphi}) \mathrm{d}\varphi_i$ (see formula 1.56). However one can use the fact that the new angles $\psi_1, \ldots, \psi_n$ are linear functions of time (since the new Hamiltonian depends only on the actions $\mathbf{J}$) and that the transformation between the old and the new variables is periodic on the cycle (as ensured by the Arnold–Liouville theorem). Therefore one must have

$$\dot{\psi}_1 = \frac{2\pi}{T} \ , \quad \dot{\psi}_l = \frac{1}{|T|} \int_0^T \frac{\partial \mathcal{H}_{\mathrm{res}}}{\partial I_l}(I_1(t), \varphi_1(t)) \mathrm{d}t \quad (l = 2, ..., n) \ , \tag{4.8}$$

where t is the time and T denotes the period of φ_1 on the cycle, with the convention that T is positive if φ_1 circulates with positive derivative, or if the motion of I_1, φ_1 evolves on cycles of clockwise libration, and it is negative otherwise. With this convention the transformation $(\mathbf{I}, \boldsymbol{\varphi}) \to (\mathbf{J}, \boldsymbol{\psi})$ is close to the identity in the circulation region, and it is close to the transformation defining the action–angle variables of a harmonic oscillator in the vicinity of the stable

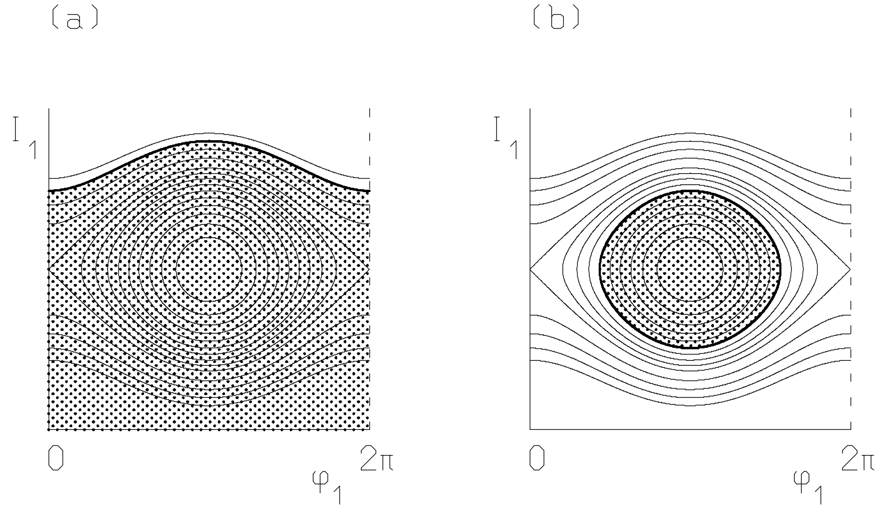


Figure 4.2: In the circulation region (panel a) the action J_1 is proportional to the area between the cycle and the $I_1 = 0$ axis, while in the libration region (panel b) it is proportional to the area enclosed by the libration cycle.

equilibrium point. The above formulæ imply that the time derivative of ψ_1 is the frequency of the motion on the cycle (frequency of libration/circulation of φ_1); the time derivatives of the other angles ψ_l ($l \neq 1$) are the averages of the time derivatives $\dot{\varphi}_l$ on the cycle. Therefore, the relationships $I_1(\psi_1)$ and $\varphi_1(\psi_1)$ on each cycle can be simply obtained from $I_1(t)$ and $\varphi_1(t)$ by substituting $\psi_1 = 2\pi t/T$. Moreover, writing $\varphi_l(t)$ as $\nu_l t + \varrho_l(t)$, with $\varrho_l(t)$ periodic with period T, the expression of φ_l as a function of ψ_1 and ψ_l ($l = 2, \ldots, n$) gives $\varphi_l = \psi_l + \varrho_l(T\psi_1/2\pi)$.

We now come to the expression of the Hamiltonian (4.1) in the resonant action–angle variables. In principle, both $\mathcal{H}_{\rm res}(\mathbf{J})$ and the remainder $\mathcal{R}(\mathbf{J}, \boldsymbol{\psi})$ can be obtained by replacing the old variables with the new ones. However, it is very hard to write the relationship between the old and the new variables in an explicit form. One has to analytically compute $I_1(t), \boldsymbol{\varphi}(t)$ and the integral (4.5). For the exact Hamiltonian of the pendulum (4.4) this requires the use of elliptic functions, but for generic resonant Hamiltonians (4.2) the analytic computation becomes practically infeasible. Nevertheless, as shown by Henrard (1990), one can efficiently compute the Hamiltonian and its derivatives in the new variables using numerical techniques. We give here only some examples of how this can be done; the reader can refer to Henrard's paper for a more exhaustive discussion.

The motion on each cycle can be numerically computed, using a numerical

integrator of the equations of motion. For each cycle, the value of the action J_1 can be computed by numerically solving the integral (4.5). This allows one to numerically determine the function $J_1(E, J_2, \ldots, J_n)$, which can be numerically inverted to compute $\mathcal{H}_{\rm res} \equiv E(J_1, J_2, \ldots, J_n)$.

The derivatives of $\mathcal{H}_{\rm res}(\mathbf{J})$ with respect to $\mathbf{J}$ can also be simply computed: the numerical integration gives the libration/circulation frequency ν_1 of φ_1 and the circulation frequencies $\nu_2, \ldots, \nu_n$ of $\varphi_2, \ldots, \varphi_n$. But these are also the frequencies of the new angles $\psi_1, \psi_2, \ldots, \psi_n$. Therefore, one has:

$$\frac{\partial \mathcal{H}_{\rm res}}{\partial J_l} = \nu_l \ , \quad l = 1, \ldots, n \ . \tag{4.9}$$

Finally, on each cycle one can expand the remainder $\mathcal{R}$ – or any other function originally written in the variables $\mathbf{I}, \boldsymbol{\varphi}$ – in Fourier series of the new angles $\boldsymbol{\psi}$ and numerically compute the coefficients of the series. This is done as follows: the numerical integration allows one to determine I_1 and $\boldsymbol{\varphi}$ on each cycle as a function of time. With the substitution $t = T\psi_1/(2\pi)$, this gives the functions $I_1(\psi_1)$, $\varphi_1(\psi_1)$ and $\varrho_l(\psi_1)$. Then one writes

$$\mathcal{R} = \mathcal{R}(I_1(\psi_1), I_2, \ldots, I_n, \varphi_1(\psi_1), \psi_2 + \varrho_2(\psi_1), \ldots, \psi_n + \varrho_n(\psi_1)) \ , \tag{4.10}$$

and numerically computes the Fourier transform of this expression. This gives the coefficients of the Fourier expansion of $\mathcal{R}(\mathbf{J}, \boldsymbol{\psi})$ with respect to $\psi_1, \ldots, \psi_n$, for each value of $\mathbf{J}$. The gradient of these coefficients with respect to $\mathbf{J}$ can then be computed by numerical differentiation.

As we have seen in the previous section, the resonant phase space is divided by the separatrix into three different dynamical regions, and the "period" on the separatrix is infinite; the cycles in the libration region and in the circulation regions have different structures. Therefore the sets of action–angle variables introduced in the libration region and in the circulation regions cannot be joined smoothly; the singularity corresponding to the separatrix cannot be eliminated. In practice, the introduction of resonant action–angle variables transforms the original integrable resonant dynamical system $\mathcal{H}_{\rm res}(\mathbf{I}, \varphi_1)$ into three distinct dynamical systems of type $\mathcal{H}_{\rm res}(\mathbf{J})$, each of which is defined on a different dynamical region. Each of these new systems is characterized by constant values of the actions $\mathbf{J}$ and linear evolutions of the angles $\boldsymbol{\psi}$, but the actions and the angles are different for the three systems. As a consequence, also the perturbation $\mathcal{R}$ in the new variables will have different expressions in the different dynamical regions. In other words, each integrable system $\mathcal{H}_{\rm res}(\mathbf{J})$ will have its own perturbation $\mathcal{R}(\mathbf{J}, \boldsymbol{\psi})$. Therefore, using resonant action–angle variables, the study of perturbed resonant dynamics breaks into three parallel studies, each focused on a different dynamical region of the main resonance.

4.3 Perturbed resonant dynamics

We now consider the effects that the remainder $\mathcal{R}$ of the optimal resonant normal form (4.1) has on the dynamics given by the integrable approximation $\mathcal{H}_{\rm res}$.

To fix notation, we will denote by *main resonance* the resonance for which the normal form has been constructed and whose dynamics is described by $\mathcal{H}_{\rm res}$; by *secondary resonances* we denote those that are present in the domain where the normal form is constructed, but whose harmonics have not been retained in $\mathcal{H}_{\rm res}$; the *leading secondary resonance* will be the secondary resonance of lowest order.

4.3.1 Size of the remainder

As explained in Section 2.4, the size of $\mathcal{R}$ is dictated by the size of the coefficient of the harmonic of the leading secondary resonance. The latter depends on the problem under consideration, and in particular on the frequencies of the angles $\varphi_2, \ldots, \varphi_n$. Speaking of "single resonance" we implicitly imply that all secondary resonances are of much larger order than the main resonance, so that $\mathcal{R}$ is small compared to $\mathcal{H}_{\rm res}$. If this is not the case, then the "single resonance" model is not appropriate, and a multiresonance problem should be considered (see Chapter 6).

There are two cases that are worthy of mention, for which we can ensure that $\mathcal{R}$ is much smaller than $\mathcal{H}_{\rm res}$, at least in the limit where the coefficient c of the main resonant harmonic (see formula 4.4) is small:

i) *Two degrees of freedom.* Recall, from Section 4.1, that the minimal size of the domain where the resonant normal form is constructed must be of order $\sqrt{c/\beta}$ in $\hat{I}_1$ to contain the entire resonant region, with c and β defined as in formula (4.4). While β is typically independent of ε, c is of order ε or smaller. On this domain $\dot{\varphi}_1 \equiv \partial\mathcal{H}_0/\partial I_1$ ranges from 0 (at $\hat{I}_1 = 0$) to $\sim \sqrt{c\beta}$ (recall that at the border of the domain $\hat{I}_1 \sim \pm\sqrt{c/\beta}$). The secondary resonances correspond to the relationships:

$$k_1\dot{\varphi}_1 + k_2\dot{\varphi}_2 = 0 \ , \tag{4.11}$$

with integer k_1, k_2 and $k_2 \neq 0$. Since $\dot{\varphi}_2 \equiv \partial\mathcal{H}_0/\partial I_2$ is of order 0 in c while $\dot{\varphi}_1$ is at most $\sim \sqrt{c\beta}$, to satisfy (4.11) one needs to have $k_1 \sim 1/\sqrt{c\beta}$ and $k_2 \sim 1$. Therefore the leading secondary resonance must be of order $\sim 1/\sqrt{c\beta}$. As a consequence, the coefficient of the corresponding harmonic and the size of the remainder $\mathcal{R}$ are $\sim \exp(-1/\sqrt{c\beta})$.

ii) *Fixed nonresonant frequencies with diophantine properties.* Assume that the frequencies of $\varphi_2, \ldots, \varphi_n$ are fixed (i.e. they don't depend on the actions) and satisfy the diophantine condition

$$\left| \sum_{l=2}^{n} k_l \dot{\varphi}_l \right| \geq \frac{\gamma}{(\sum_{l=2}^{n} |k_l|)^\tau} \ , \quad (k_2, \ldots, k_n) \in \mathbf{Z}^{n-1} \tag{4.12}$$

for some positive γ and τ. The secondary resonances then correspond to the relationships:

$$k_1 \dot{\varphi}_1 + \sum_{l=2}^{n} k_l \dot{\varphi}_l = 0 \tag{4.13}$$

with $k_2, \ldots, k_n$ not all simultaneously equal to zero. Because $\dot{\varphi}_1$ is at most $\sim \sqrt{c\beta}$, and because of (4.12), the relations (4.13) can be satisfied only if $\sum_{l=2}^{n} |k_l| \sim 1/\sqrt{c\beta}^{1/\tau}$. As a consequence the leading secondary resonance must be of order $1/\sqrt{c\beta}^{1/\tau}$. Therefore, the coefficient of the corresponding harmonic and the size of the remainder $\mathcal{R}$ are $\sim \exp(-1/\sqrt{c\beta}^{1/\tau})$.

4.3.2 Resonant invariant tori

As shown in the previous section, by introducing resonant action–angle variables in each dynamical region, the Hamiltonian (4.1) can be written as $\mathcal{H}_{\rm res}(\mathbf{J}) + \mathcal{R}(\mathbf{J}, \boldsymbol{\psi})$. Under the assumption that $\mathcal{H}_{\rm res}$ is nondegenerate, the KAM theorem can therefore be applied. As seen in Chapter 3, if the perturbation $\mathcal{R}$ is sufficiently small, there must exist a set of large measure of invariant tori. These tori must be "close" to the invariant tori for the integrable approximation $\mathcal{H}_{\rm res}$, i.e. close to the tori defined by $\mathbf{J} =$constant. Therefore, in the variables I_1, φ_1 the invariant tori of the perturbed resonance will be close to the invariant curves of the pendulum (see Fig. 4.1). In particular, in the circulation regions the invariant tori have the same structure as those already discussed in the previous chapter: the angles $\varphi_1, \ldots, \varphi_n$ range from 0 to 2π, and the actions $\mathbf{I}$ are periodic functions of the angles. In the libration region, conversely, the motion on the invariant torus will result in the circulation of the angles $\varphi_2, \ldots, \varphi_n$ and in the oscillation of φ_1 on a bounded interval; we will call this a *librational* or *resonant* invariant torus.

As seen in Chapter 3, invariant tori can be well visualized in two-degree of freedom problems, using Poincaré sections. Figure 4.3 gives an example using the standard map (see formula 3.13). In this case, the initial conditions have been chosen on the $q_1 = \pi$ axis, in order to explore also the dynamics in the region associated to the resonance $\dot{q}_1 = 0$ that appeared empty in Fig. 3.3.

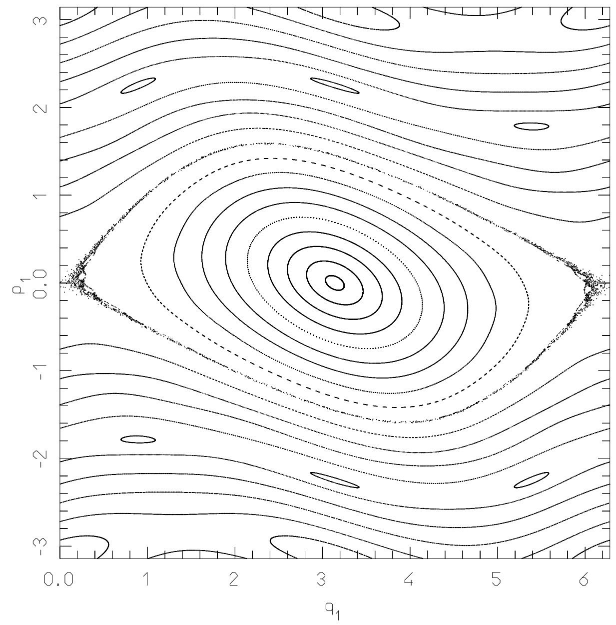


Figure 4.3: The phase portrait of the standard map for $\varepsilon = 0.6$.

In Fig. 4.3 (drawn for the same value of ε as in Fig. 3.3b) the resonant region appears to be filled by closed curves: these are the images of the librational invariant tori on the Poincaré section. Outside of the resonant region, we find again the invariant tori already seen in Fig. 3.3b. The chains of islands visible above and below the main resonant region are the images of the librational invariant tori associated to other resonances, of type $k_1\dot{q}_1 + k_2\dot{q}_2 = 0$. Notice that for these resonances the Poincaré section on q_2 reproduces the typical pendulum-like portrait k_1 times on the $[0, 2\pi]$ interval of q_1. Recalling that in the standard map (3.13) $\dot{q}_2 = 2\pi$ and $\dot{q}_1 \sim p_1$, it follows that the chain of three islands at $p_1 \sim 2\pi/3$ corresponds to the resonance $3\dot{q}_1 - \dot{q}_2 = 0$ and the chain of two islands at $p_1 \sim \pi$ corresponds to the resonance $2\dot{q}_1 - \dot{q}_2 = 0$. If we zoomed the image around one of these islands, we would reveal a structure very similar to that of the big island at the center of the panel. Islands of course could also appear among librational invariant tori, related to secondary resonances between the libration period of the critical angle of the main resonance and the circulation period of q_2 (these are the resonances between the angles ψ_1 and ψ_2 introduced in the libration region for the integrable approximation $\mathcal{H}_{\rm res}$ of the main resonance). In Fig. 4.3 these islands are not visible, because they are too small. However an example can be seen in Fig. 3.3d, where a chain of eight islands surrounds the main resonant region.

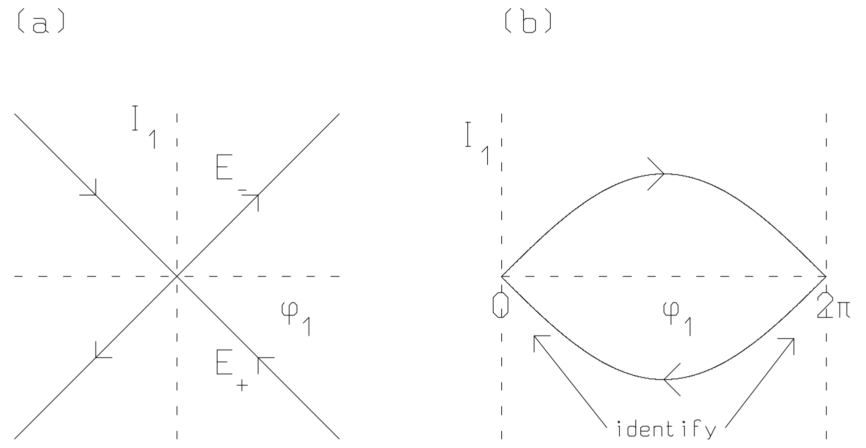


Figure 4.4: Left: stable and unstable manifolds of the hyperbolic equilibrium point in the local linear approximation of $\mathcal{H}_{\mathrm{res}}$. Right: the stable and unstable manifolds for the pendulum.

The KAM theorem cannot be applied close to the separatrices of the integrable approximation $\mathcal{H}_{\mathrm{res}}$, because the latter, when written in resonant action–angle variables, loses its regularity properties in correspondence to the separatrix. On the other hand, Fig. 4.3 shows that there are no invariant curves at the boundary between the libration and the circulation regions. The separatrix, typical of the pendulum model, has disappeared and its place has been taken by a *chaotic region*, which appears on the Poincaré section as a collection of scattered points. To understand the origin of the chaotic region, we have to analyze in detail what happens to the separatrix of the integrable approximation $\mathcal{H}_{\mathrm{res}}$ in the presence of an arbitrarily small perturbation $\mathcal{R}$.

4.3.3 Splitting of separatrices

The unstable equilibrium point of the pendulum at $\hat{I}_1 = \varphi_1 = 0$ has a hyperbolic structure. The second-order Taylor expansion of the Hamiltonian (4.4) around this point gives

$$\mathcal{H}_{\mathrm{res}} = \frac{\beta}{2}\hat{I}_1^2 - \frac{c}{2}\varphi_1^2 \tag{4.14}$$

whose trajectories are hyperbolæ having for asymptotes the lines E_-, of equation $\hat{I}_1 = \sqrt{c/\beta}\varphi_1$, and E_+, of equation $\hat{I}_1 = -\sqrt{c/\beta}\varphi_1$ (Fig. 4.4a). On E_- the motion is $\varphi_1(t) = \varphi_1(0)\exp(\sqrt{c\beta}t)$, $\hat{I}_1(t) = \sqrt{c/\beta}\varphi_1(0)\exp(\sqrt{c\beta}t)$, which tends to the equilibrium point for $t \to -\infty$: we will call this line

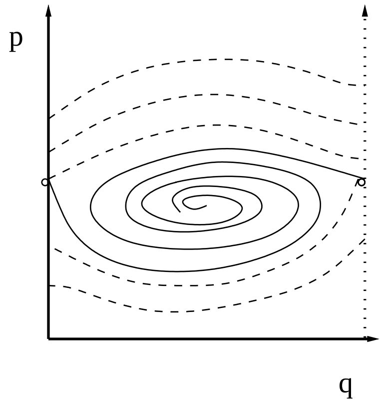


Figure 4.5: If the unstable (dashed curves) and stable (solid curves) manifolds did not intersect, two of them should spiral towards the center of the resonance. But this would violate the property of volume conservation of Hamiltonian flow. Therefore stable and unstable manifolds must intersect.

the *unstable manifold*; on E_+ the motion is $\varphi_1(t) = \varphi_1(0)\exp(-\sqrt{c\beta}t)$, $\hat{I}_1(t) = -\sqrt{c/\beta}\varphi_1(0)\exp(-\sqrt{c\beta}t)$, which tends to the equilibrium point for $t \to +\infty$: we will call this line the *stable manifold*. Taking into account the higher-order terms in $\hat{I}_1$ and φ_1 in the Taylor expansion, the stable and the unstable manifolds are no longer straight lines, but more generally they are curves in the $\hat{I}_1, \varphi_1$ plane. For the pendulum the stable manifold of the equilibrium point at $\varphi_1 = 0$ coincides with the unstable manifold of the equilibrium point at $\varphi_1 = 2\pi$ and vice versa (Fig. 4.4b). Stable and unstable manifolds constitute what we have called the *separatrix* in Section 4.1.

We now consider an additional uncoupled degree of freedom and look at the pendulum dynamics in extended phase space. To fix ideas, consider the integrable Hamiltonian

$$\mathcal{H}(\hat{I}_1, I_2, \varphi_1) = I_2 + \mathcal{H}_{\rm res}(\hat{I}_1, \varphi_1) \ , \tag{4.15}$$

where $\mathcal{H}_{\rm res}$ is a one-degree of freedom pendulum-like Hamiltonian as in (4.4), and the action–angle variables I_2, φ_2 represent an additional uncoupled degree of freedom. The dynamics in the $\hat{I}_1, \varphi_1$ plane is of course unchanged, so that the Poincaré section of (4.15) on φ_2 gives a portrait identical to the one of the pendulum (Fig. 4.1). Introduce now an arbitrarily small coupling between the two degrees of freedom, namely consider the generically nonintegrable Hamiltonian

$$\mathcal{H} = I_2 + \mathcal{H}_{\rm res}(\hat{I}_1, \varphi_1) + \mathcal{R}(\varphi_1, \varphi_2) \ , \tag{4.16}$$

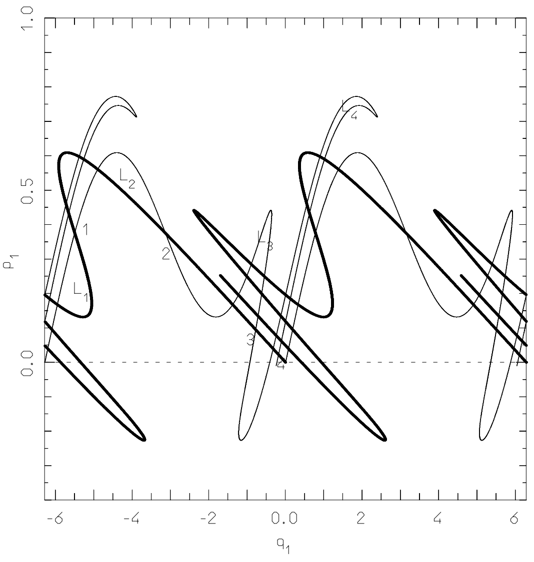


Figure 4.6: The transversal intersections between the stable (bold curve) and unstable (light curve) manifolds of the (0,0) unstable equilibrium point of the standard map. Only half of the picture (that produced by the upper manifolds) is plotted, for simplicity. The parameter ε is equal to 2.5.

with $\mathcal{H}_{\rm res}$ still as in (4.4). It can be proven (Poincaré, 1892) that the Poincaré section of the new system still shows a hyperbolic fixed point close to $\hat{I}_1 = \varphi_1 = 0$ (as well as a stable fixed point close to $\hat{I}_1 = 0, \varphi_1 = \pi$). Moreover, the stable and unstable manifolds still have to intersect, because otherwise one of the manifolds should spiral towards the center of the resonance as sketched in Fig. 4.5, in manifest violation of the conservation of volume that is characteristic of Hamiltonian systems (see Section 1.7). However, there is no reason to expect that the stable and unstable manifolds still coincide, as in the pendulum case. Generically, they will intersect transversally. This is known as the phenomenon of *homoclinic intersection* or *separatrix splitting* (see Poincaré, 1892; Arnold and Avez, 1968).

Figure 4.6 shows for the standard map (3.13) the upper unstable (light curve) and stable (solid curve) manifolds of the hyperbolic point at $q_1 = 0$ on the interval $[-2\pi, 2\pi]$ (the reader can find in Simó, 1990, the explanation of the standard algorithm for the numerical computation of the manifolds). As one sees, the two manifolds intersect transversally; each point of intersection is called a *homoclinic point.* As the figure suggests, there is an infinite number of homoclinic points: in fact, choose one homoclinic point as the initial condition of one orbit, say the one indicated by the number 1 in Fig. 4.6. Since the initial

condition is on both the stable and the unstable manifold, the whole orbit is, by definition, on both manifolds. Moreover, the successive images of this orbit on the Poincaré section (the points indicated by 2, 3, 4, ...) cannot coincide, because otherwise the orbit would be periodic and would not tend to the hyperbolic fixed point; this proves that the orbit must leave infinitely many distinct points on the Poincaré sections which, by definition, are all homoclinic points.

As Fig. 4.6 also shows, the distance between two successive homoclinic points decreases, because the orbit "slows down" exponentially approaching the hyperbolic fixed point. Now pay attention to the area enclosed by the stable and the unstable manifolds, labeled by L_1 in Fig. 4.6. We call this area a *lobe*. By a continuity argument, one can prove that this lobe is mapped onto L_2, and, successively, onto L_i, $i = 3, \ldots, +\infty$, which are thus all successive iterations of the *same* lobe. The area of each L_i must be the same, because the Hamiltonian flow preserves areas; thus, because the length of the bases of the L_is decreases exponentially with i, the length of their "heights" must increase exponentially. The picture gets complicated: different lobes must intersect, as shown in the figure, generating homoclinic intersections of second order between the stable and the unstable manifolds. The result is what is usually called the *homoclinic tangle*. The region densely criss-crossed by the stable and unstable manifolds is called the *chaotic region*. The dynamics in the chaotic region is equivalent to that of the well-known *Smale horseshoe*, (Smale, 1963, 1980; see also Wiggins, 1988). Arbitrary close initial conditions belong to different parts of the homoclinic tangle (i.e. to different lobes) and therefore their dynamical evolutions are completely different. The distance between their orbits diverges exponentially with time. As shown in Fig. 4.3, the Poincaré section of an orbit in the chaotic region gives a collection of scattered points whose distribution misses any apparent regularity; the resonant angle for chaotic trajectories alternates between libration and circulation in an apparently random fashion.

4.3.4 Size of the chaotic region

The size of the chaotic region depends on the "distance" between the stable and the unstable manifolds. As suggested by Fig. 4.6, the distance between the manifolds can be measured on the Poincaré section as the difference between the values of the action on the two manifolds, for a given value of the angle. If the stable and the unstable manifolds coincide, as in the case of the integrable resonance, the distance is zero for all values of the angle; conversely, if the separatrix splits, the distance is zero only in correspondence to the homoclinic points.

In principle, the distance between the manifolds can be measured using the so-called Poincaré–Melnikov integral (Poincaré, 1892; Melnikov, 1983; see also Arnold *et al.*, 1988). The idea is quite simple. Consider again the Hamiltonian (4.16): on the Poincaré section, once φ_1 is fixed, there is a local one-to-one correspondence between $\hat{I}_1$ and the value of $\mathcal{H}_{\rm res}$. Therefore, the distance between two points on the Poincaré section which have the same value of φ_1 can be measured as the difference between the respective values of $\mathcal{H}_{\rm res}$. Denote now by $\mathcal{H}^+_{\rm res}(\varphi_1^0)$ the value of $\mathcal{H}_{\rm res}$ on the stable manifold when it first intersects the axis $\varphi_1 = \varphi_1^0$. Analogously, denote by $\mathcal{H}^-_{\rm res}(\varphi_1^0)$ the corresponding value on the unstable manifold. The distance between the manifolds can then be measured as $\Delta\mathcal{H}_{\rm res} = \mathcal{H}^+_{\rm res}(\varphi_1^0) - \mathcal{H}^-_{\rm res}(\varphi_1^0)$. Using (1.47) one has

$$\begin{aligned}\mathcal{H}^+_{\rm res}(\varphi_1^0) &= \mathcal{H}_{\rm res}(+\infty) - \int_0^{+\infty} \{\mathcal{H}_{\rm res}, \mathcal{H}\}{\rm d}t \\ \mathcal{H}^-_{\rm res}(\varphi_1^0) &= \mathcal{H}_{\rm res}(-\infty) + \int_{-\infty}^{0} \{\mathcal{H}_{\rm res}, \mathcal{H}\}{\rm d}t \end{aligned} \tag{4.17}$$

where the first integral is computed along the stable manifold, the second integral along the unstable manifold, and $\mathcal{H}_{\rm res}(+\infty) = \mathcal{H}_{\rm res}(-\infty)$ are the values of $\mathcal{H}_{\rm res}$ on the hyperbolic fixed point (reached respectively at $t = +\infty$ on the stable manifold and $t = -\infty$ on the unstable manifold). The origin of time has been arbitrarily set equal to 0, since the function $\{\mathcal{H}_{\rm res}, \mathcal{H}\}$ does not explicitly depend on time; the integrals depend only on the values of the coordinates at time 0.

The integrals in (4.17) cannot be computed, because one does not explicitly know the solution of the equations of motion on the stable and the unstable manifolds. One notes, however, that the manifolds of $\mathcal{H}$ differ from the manifolds of $\mathcal{H}_{\rm res}$ by a quantity proportional to the size of $\mathcal{R}$, say μ. On the other hand, also the Poisson bracket $\{\mathcal{H}_{\rm res}, \mathcal{H}\}$ is of order μ. Therefore computing the integrals in (4.17) along the manifolds of $\mathcal{H}_{\rm res}$ one makes an approximation not worse than order μ^2. The advantage is that the equations of motion on the manifolds of $\mathcal{H}_{\rm res}$ are known, because $\mathcal{H}_{\rm res}$ is integrable, so that in principle the integrals can be computed.

With this approximation one gets

$$\Delta\mathcal{H}_{\rm res} = -\int_{-\infty}^{+\infty} \{\mathcal{H}_{\rm res}, \mathcal{H}\}{\rm d}t + O(\mu^2)\ , \tag{4.18}$$

the integral being computed on the separatrix of $\mathcal{H}_{\rm res}$. The integral in (4.18) is usually called the *Poincaré–Melnikov integral*, and it depends only on the chosen value φ_1^0 at $t = 0$.

As an example, let us compute the Poincaré–Melnikov integral for the model (4.16), with $R(\varphi_1, \varphi_2) = \mu\varphi_1 \sin\varphi_2$ and $\mathcal{H}_{\rm res}$ as in (4.4). One has

$\{\mathcal{H}_{\rm res}, \mathcal{H}\} = -\mu\beta\hat{I}_1 \sin\varphi_2$. The solution of the equations of motion on the separatrix of $\mathcal{H}_{\rm res}$ in (4.4) are

$$\hat{I}_1(t) = \pm 2\frac{\sqrt{c/\beta}}{\cosh\tau} \qquad \varphi_1(t) = \pm \mathrm{arc\,cot}(-\sinh\tau) \ , \tag{4.19}$$

where $\tau = \sqrt{c\beta}(t - s_0)$. The initial value of φ_1 and $\hat{I}_1$ at $t = 0$ is therefore determined by the parameter s_0. In addition, from the equations given by Hamiltonian (4.15) one has

$$\varphi_2(t) = t + \varphi_2(0) \tag{4.20}$$

where $\varphi_2(0)$ is the initial phase (assume 0 for simplicity). The Melnikhov integral is therefore

$$\begin{aligned} M(s_0) &= \mu\beta \int_{-\infty}^{+\infty} \hat{I}_1(t) \sin\varphi_2(t) \mathrm{d}t \\ &= \mu\beta \int_{-\infty}^{+\infty} \frac{2\sqrt{c/\beta}}{\cosh(\sqrt{c\beta}(t - s_0))} \sin t \mathrm{d}t \\ &= -\frac{2\pi\mu}{\cosh\frac{\pi}{2\sqrt{c\beta}}} \sin s_0 \ . \end{aligned} \tag{4.21}$$

As one sees, $M(s_0)$ is alternatively zero, negative or positive, depending on s_0. This proves that i) there exist homoclinic intersections, which occur for the values φ_1^0 of φ_1 given by $s_0 = k\pi$ ($k \in \mathbf{Z}$) in (4.19) and ii) the intersections are transversal because the distance between the manifolds is zero only for discrete values of s_0 (i.e. of φ_1^0).

The computation of the Poincaré–Melnikov integral, although feasible in principle, is very difficult in practice, apart from simple cases where the motion on the separatrix of $\mathcal{H}_{\rm res}$ and the Poisson bracket $\{\mathcal{H}_{\rm res}, \mathcal{H}\}$ have a simple expression (see Delshams *et al.*, 1999, for a review). Moreover, in several cases, the result of the Poincaré–Melnikov integral is small compared to μ^2, so that the neglected term in (4.18), which is in principle as large as μ^2, could dominate the result. In these cases, it would be necessary to look for more complicated formulæ for $\Delta\mathcal{H}_{\rm res}$ that are exact up to higher order in μ. As a consequence, in most of the cases the rigorous analytic computation of the amplitude of separatrix splitting is an open problem.

However, heuristic arguments and numerical experiments (see Morbidelli and Giorgilli, 1997) show that the amplitude of the splitting is in most cases related to the coefficient of the leading secondary resonance (see the top of Section 4.3 for its definition), i.e. to the size of the remainder $\mathcal{R}$ of the optimal

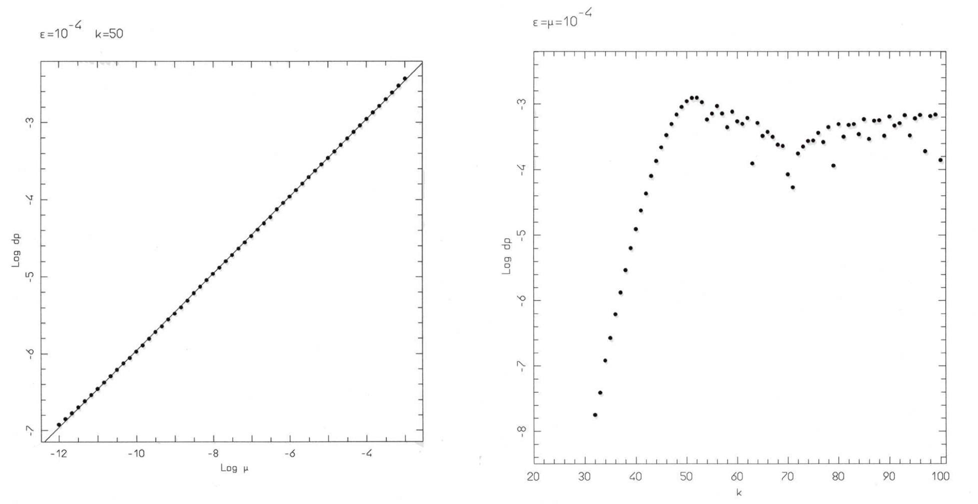


Figure 4.7: Panel a (left): the amplitude of the splitting in (4.22) as a function of μ. Panel b (right): the same as a function of k. Reprinted from Figs. 4 and 5 of Morbidelli and Giorgilli (1997), with permission from Elsevier Science.

normal form. Let us consider as an example the Hamiltonian $\mathcal{H}(p,q,t) = \mathcal{H}_{\rm res}(p,q) + \mathcal{R}(p,q,t)$ with:

$$\mathcal{H}_{\rm res}(p,q) = \frac{1}{2}p^2 - \varepsilon\cos q\,;\ \mathcal{R}(p,q,t) = -\frac{\varepsilon\mu}{2}\{\cos[(k+1)q-t]+\cos[(k-1)q-t]\}\,. \tag{4.22}$$

Choosing the integer k close to $1/(2\sqrt{\varepsilon})$ the harmonics in $\mathcal{R}$ are resonant at the border of the domain $|p| < 2\sqrt{\varepsilon}$ that is spanned by the separatrices of the pendulum Hamiltonian $\mathcal{H}_{\rm res}$, and thus they correspond to secondary resonances. Therefore, in this case the Hamiltonian (4.22) can be considered as the paradigm of the optimal normal form of a perturbed pendulum problem. Figure 4.7a shows the amplitude of the separatrix splitting (measured numerically as the distance Δp between the stable and the unstable manifolds at $q = \pi$) as a function of μ; ε and k are fixed equal to 10^{-4} and 50 respectively. The amplitude of the splitting perfectly scales as $\sqrt{\mu}$ over nine orders of magnitude in μ. The power of μ depends on the value of q for which the amplitude of the splitting is computed: at $q = 0$, the scaling would have become proportional to μ. Figure 4.7b shows the amplitude of the splitting as a function of k, for fixed $\varepsilon = \mu = 10^{-4}$. For $k \geq 1/(2\sqrt{\varepsilon}) = 50$ the harmonics in $\mathcal{R}$ correspond to secondary resonances (since they are located at $p \sim 1/k$, i.e. inside the domain $|p| \leq 2\sqrt{\varepsilon}$ spanned by the separatrices), and the amplitude of the splitting is

constant within an order of magnitude. For $k < 50$ the harmonics in $\mathcal{R}$ do not correspond to secondary resonances so that (4.22) cannot be considered as an optimal normal form. Using the normalization algorithm explained in Chapter 2, one could proceed to the elimination of nonresonant harmonics, until the harmonics related to secondary resonances are generated; this would happen only at order $|k - 50|$ in $\varepsilon\mu$, so that their coefficients would be of size $O((\varepsilon\mu)^{|k-50|})$. Figure 4.7b shows in fact that the amplitude of the splitting drops exponentially with k for $k < 50$.

This example shows that the "amount" of nonintegrability, that is the size of the remainder of the optimal normal form, is intimately related to the "amount" of chaos, that is the amplitude of separatrix splitting. The appearance of chaos is a criterion of nonintegrability of a given Hamiltonian system (a rigorous result known as the Ziglin theorem; Ziglin, 1980; see also Arnold *et al.*, 1988). If it is very difficult to rigorously prove the existence of chaos (since one must solve the Poincaré–Melnikov integral), several powerful numerical tools – nowadays largely in use in Celestial Mechanics – allow to unequivocally reveal chaotic behavior.

Chapter 5

NUMERICAL TOOLS FOR THE DETECTION OF CHAOS

5.1 Monitoring the time evolution in phase space

How is it possible to show that a given initial condition generates chaotic evolution? How is it possible to show that a given orbit lies on an invariant torus? These questions become important at this point.

We have seen in the previous chapter that the computation of the Poincaré section allows one to unequivocally distinguish regular from chaotic orbits. However, the Poincaré section is useful only for two-degree of freedom systems. In more degrees of freedom, in fact, the "surface" of the section has more than two dimensions, so that it is not possible to display the Poincaré section in a graphical way.

The time evolution of the canonical variables $\mathbf{p}, \mathbf{q}$ might also reveal, at a glance, the dynamical character of the orbit. For the simple Hamiltonian $\mathcal{H}(p, q, t) = p^2/2 + \cos q + 1/4[\cos(q + 5t) + \cos(q - 5t)]$, Fig. 5.1 shows the time evolution of the action p and the angle q for a chaotic orbit (panel a) and a regular orbit (panel b). The difference between the two evolutions is evident. For the chaotic orbit, the angle q alternates among large amplitude oscillations around π, circulations from 0 to 2π with positive derivative, and circulations from 2π to 0 with negative derivative. Correspondingly, the conjugate action p oscillates around $p \sim 1$ when q circulates with positive derivative, oscillates around $p \sim -1$ when q circulates with negative derivative, and oscillates around $p \sim 0$ – with amplitude that is about twice as large – when q librates around π. It is evident that these three behaviors correspond to the three

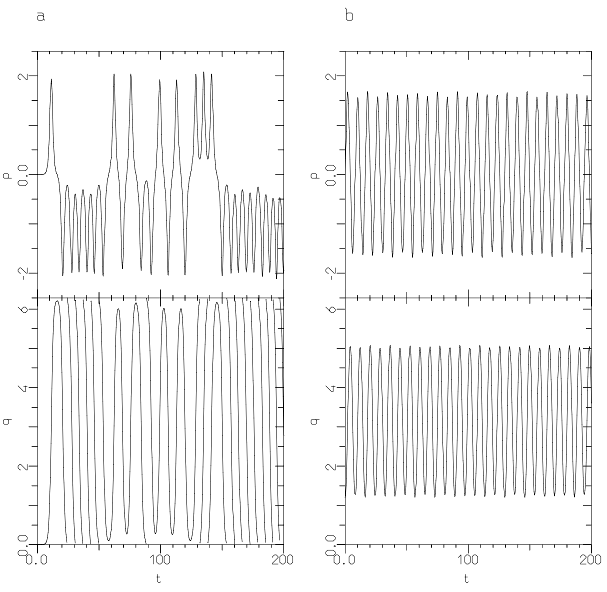


Figure 5.1: Time evolution for a chaotic (panel a) and a regular orbit (panel b) of a perturbed pendulum. See text for comments.

dynamical regimes close to the separatrix of the integrable resonant approximation $\mathcal{H}_{\rm res} = p^2/2 + \cos q$; the characteristic of chaotic orbits is precisely the alternation among the different regimes. The transitions from one to another regime repeat without any apparent regularity. Conversely, a regular orbit (panel b) does not alternate among the different dynamical regimes; p and q oscillate in a regular, quasi-periodic way.

However, the situation is not so simple when one considers more complicated dynamical systems, with larger number n of degrees of freedom. In these cases, due to the coupling among the degrees of freedom, also for regular orbits the actions and the angles show a complicated time dependence, characterized by n independent frequencies. Therefore, chaotic orbits may not differ too much from regular ones, and it becomes impossible to visually distinguish them. As an example, Fig. 5.2 shows by a solid line the time evolution of the eccentricity of the orbit of the Earth – which is chaotic (see Chapter 7) – and by a dotted line the evolution of a similar but regular orbit, issued from an integrable analytical approximation.

The point is that for a quasi-integrable Hamiltonian $\mathcal{H}_0(\mathbf{p}) + \varepsilon\mathcal{H}_1(\mathbf{p},\mathbf{q})$ chaos may be raised by the separatrix splitting of a resonance whose corresponding harmonic has a quite small coefficient, say for example of size ε^4.

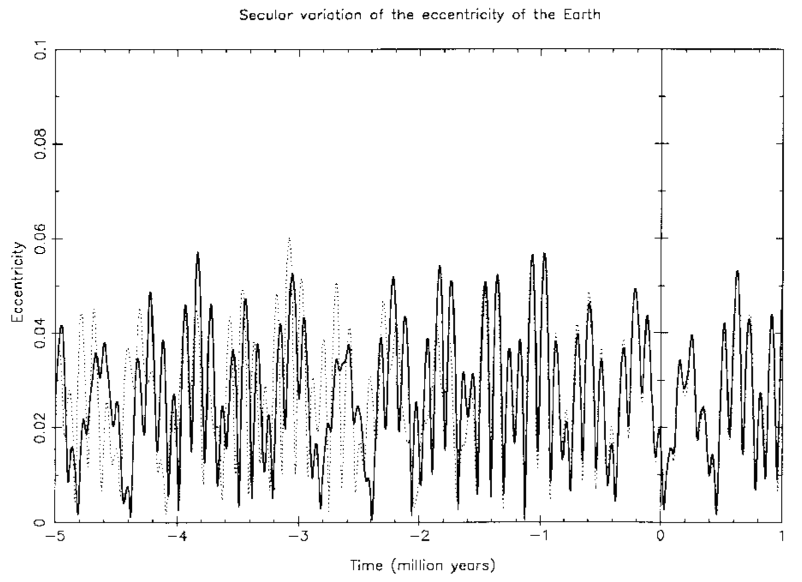


Figure 5.2: The time evolution of the eccentricity of the Earth according to numerical integration (solid curve) and to an analytic model (dotted curve). It is difficult to say which evolution is regular and which is chaotic. Reprinted from Fig. 1 of Laskar (1988), with permission from Astronomy and Astrophysics.

In this case, one can introduce new variables $\mathbf{p}^1, \mathbf{q}^1$ in order to transform the Hamiltonian into the resonant normal form of type $\mathcal{H}_{\rm res} = \mathcal{H}_0'(\mathbf{p}^1) + \varepsilon^4\mathcal{H}_{\rm norm}(\mathbf{p}^1, \mathbf{q}^1) + \mathcal{R}(\mathbf{p}^1, \mathbf{q}^1)$, with $\mathcal{R}$ smaller than ε^4. Then, visual inspection of the time evolution of the variables $\mathbf{p}^1, \mathbf{q}^1$ would easily reveal whether the orbit is regular or chaotic, as in the example of Fig. 5.1. The size of the regular/chaotic oscillations of $\mathbf{p}^1$ are of order ε^2, i.e. the square root of the size of the coefficient of the main resonant harmonic, as shown in the previous chapter. However, the transformation from the original variables $\mathbf{p}, \mathbf{q}$ to the new variables $\mathbf{p}^1, \mathbf{q}^1$, is of order ε and periodic in the angles $\mathbf{q}^1$ (see Chapter 2). Therefore, the time evolution of $\mathbf{p}, \mathbf{q}$ shows quasi-periodic oscillations with amplitude ε, completely hiding the chaotic oscillations, whose amplitudes are only ε^2.

In conclusion, one cannot in general rely on the visual analysis of the time evolution of the variables of the system to conclude anything about the regular/chaotic character of the orbits. It is therefore necessary to look for more precise and mathematically founded tools.

In the following we detail on Lyapunov exponents, frequency analysis and on their surrogates, which are by far the most frequently used tools for the investigation of Solar System dynamics. Conversely, we will not to elaborate about tools like the KS–entropy (Krylov, 1950; Kolmogorov, 1959; Sinai, 1959; see also Lichtenberg and Lieberman, 1983) and the Ricci curvature (Gurzadyan and Kocharyan, 1987; El-Zant 1997; El-Zant and Gurzadyan, 1998), because they are mostly used to characterize globally chaotic systems with ergodic properties – like the N-body problems with very large N – typical of galactic dynamics. The reader concerned by the characterization of chaos in these systems can find several interesting chapters in Gurzadyan and Ruffini (2000).

5.2 Lyapunov exponents

As we have seen in the previous chapter, in a chaotic region arbitrarily close initial conditions lead to evolutions that diverge exponentially fast with time. The *Maximum Lyapunov Exponent* (MLE) characterizes the rate of such exponential divergence. A detailed treatment of the theory of Lyapunov exponents goes beyond the scopes of this chapter; readers can instead refer to the papers by Lyapunov (1907), Oseledec (1968) and Benettin *et al.* (1976, 1980) and also to Lichtenberg and Lieberman (1983). Here we just outline the basic ideas and properties that are important for applications in Celestial Mechanics.

Consider two orbits, with initial conditions $\mathbf{p}^{(1)}(0), \mathbf{q}^{(1)}(0)$ and $\mathbf{p}^{(2)}(0), \mathbf{q}^{(2)}(0)$; to measure the relative divergence of their time evolutions $(\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t))$ and $(\mathbf{p}^{(2)}(t), \mathbf{q}^{(2)}(t))$ one cannot just measure their Euclidean distance $\left\|\mathbf{p}^{(1)}(t) - \mathbf{p}^{(2)}(t)\right\| + \left\|\mathbf{q}^{(1)}(t) - \mathbf{q}^{(2)}(t)\right\|$ as a function of time. In fact, in the case of bounded motion, the relative distance cannot grow indefinitely. To fix ideas, consider two trajectories on the surface of a sphere: it is evident that, despite their initial rate of relative divergence, their mutual distance can never exceed the diameter of the sphere. The exponential divergence can be measured only *locally*. It is therefore convenient to linearize the equations of relative motion.

Denoting $\delta\mathbf{p}(t) = \mathbf{p}^{(2)}(t) - \mathbf{p}^{(1)}(t)$ and $\delta\mathbf{q}(t) = \mathbf{q}^{(2)}(t) - \mathbf{q}^{(1)}(t)$, the linearized equation of relative motion can be written:

$$\begin{aligned} \dot{\delta p_i} &= -\frac{\partial^2\mathcal{H}}{\partial p_j \partial q_i}(\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t))\delta p_j - \frac{\partial^2\mathcal{H}}{\partial q_j \partial q_i}(\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t))\delta q_j \\ \dot{\delta q_i} &= \frac{\partial^2\mathcal{H}}{\partial p_j \partial p_i}(\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t))\delta p_j + \frac{\partial^2\mathcal{H}}{\partial p_i \partial q_j}(\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t))\delta q_j \end{aligned} \tag{5.1}$$

where $\mathcal{H}(\mathbf{p}, \mathbf{q})$ is the Hamiltonian of the system of which $\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t)$ and

$\mathbf{p}^{(2)}(t), \mathbf{q}^{(2)}(t)$ are trajectories. In the above formula $\delta p_i, \delta q_i, p_i, q_i$ denote the components of $\delta\mathbf{p}, \delta\mathbf{q}, \mathbf{p}, \mathbf{q}$ and summation over the index j is implicit.

In principle, equations (5.1) should be followed only for a time such that $\delta\mathbf{p}(t)$ and $\delta\mathbf{q}(t)$ stay small enough to consider the linearized equations as a good approximation of the real relative motion of the two trajectories. However, on a limited time interval, the motion of $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$ and its growth rate depend of course on the initial condition $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$ and on the considered time interval. This would be quite inconvenient to characterize the dynamical properties of the orbit $\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t)$. Therefore, forgetting about the original meaning of $\delta\mathbf{p}, \delta\mathbf{q}$, the maximum Lyapunov exponent (MLE) is defined as:

$$\mathcal{L} = \lim_{t\to+\infty} \frac{1}{t} \ln \frac{\|\delta\mathbf{p}(t), \delta\mathbf{q}(t)\|}{\|\delta\mathbf{p}(0), \delta\mathbf{q}(0)\|} \ , \tag{5.2}$$

where $\|\delta\mathbf{p}(t), \delta\mathbf{q}(t)\|$ denotes the Euclidean norm of the vector of components $(\delta p_1(t), \ldots, \delta p_n(t), \delta q_1(t), \ldots, \delta q_n(t))$. It can be proven that $\mathcal{L}$ is the same for most initial choices of $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$ (i.e. the volume of the set of $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$ that lead to a different value of $\mathcal{L}$ is zero), and that its value is also independent of the choice of canonical variables $\mathbf{p}, \mathbf{q}$, used to write the equations of motion (5.1) (Oseledec, 1968). Thanks to these properties, one can conclude that the MLE characterizes the orbit $\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t)$. In principle, each orbit of a dynamical system has its own MLE.

In the case where the trajectory $\mathbf{p}^{(1)}(t), \mathbf{q}^{(1)}(t)$ is a fixed point, namely $\mathbf{p}^{(1)}(t) = \mathbf{p}^{(1)}(0)$ and $\mathbf{q}^{(1)}(t) = \mathbf{q}^{(1)}(0)$, the equations (5.1) become linear with constant coefficients, and are therefore integrable. Denote by $\lambda_1, \ldots, \lambda_{2n}$ the (possibly complex) eigenvalues of the matrix of the second-order derivatives of $\mathcal{H}$ in $\mathbf{p}^{(1)}(0), \mathbf{q}^{(1)}(0)$. Because the Hamiltonian system preserves volume, if λ is an eigenvalue then $1/\lambda$ is also an eigenvalue. Denote by $\bar\lambda$ the eigenvalue such that $|\bar\lambda| = \max_{i=1,\ldots,2n} |\lambda_i|$; then, either $|\bar\lambda| > 1$ or $|\bar\lambda| = 1$. In the first case, the linearized dynamics (5.1) has at least one hyperbolic component, namely there exist at least two conjugate variables whose behavior is like in Fig. 4.4a. Therefore, for all initial conditions $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$ – except those that have null projection on the eigenvector associated to $\bar\lambda$ (a set of zero measure) – $\|\delta\mathbf{p}(t), \delta\mathbf{q}(t)\|$ will grow for $t \to +\infty$ as $\exp \bar\lambda t$. As a consequence, from (5.2), one has $\mathcal{L} = \bar\lambda$. Conversely, in the case $|\bar\lambda| = 1$ the linearized dynamics is elliptic, and the evolution $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$ is bounded, so that $\mathcal{L} = 0$.

The MLE of the orbits on KAM tori is also 0. This can be seen as follows. In the neighborhood of the torus one can introduce new action–angle variables $\mathbf{I}, \boldsymbol{\varphi}$ to transform the Hamiltonian into the Kolmogorov normal form (3.1). In these variables, the equations (5.1) for an orbit on the torus become:

$$\dot{\delta\mathbf{I}} = \mathbf{0} \ , \quad \dot{\delta\boldsymbol{\varphi}} = C\delta\mathbf{I} + F(\boldsymbol{\varphi}(t))\delta\mathbf{I} \ , \tag{5.3}$$

where $\boldsymbol{\varphi}(t)$ is the vector of the angles on the torus, that evolves linearly with t; the matrix C is constant, while the matrix F is periodic in $\boldsymbol{\varphi}$. As a consequence, the motion relative to the orbit of the torus is given by $\delta\mathbf{I} =$ constant and $\delta\boldsymbol{\varphi} = \boldsymbol{\alpha}t + \boldsymbol{\beta}(t)$ with constant $\boldsymbol{\alpha}$ and periodic function $\boldsymbol{\beta}(t)$. Therefore there is no exponential divergence of the orbits in the vicinity of the torus, which, according to (5.2), gives $\mathcal{L} = 0$.

Because chaotic orbits have $\mathcal{L} > 0$ while orbits on KAM tori have $\mathcal{L} = 0$, a positive MLE is usually considered as an indicator of chaos. However, a positive MLE does not really imply that the considered orbit is chaotic, if we intend the word *chaos* as a synonym for *irregular and unpredictable behavior.* Strictly speaking, a positive Lyapunov exponent just indicates the presence of *local hyperbolicity* in the neighborhood of the orbit. For instance, the hyperbolic fixed point of a pendulum has positive MLE, despite its evolution being highly regular and predictable! However, remember that in the pendulum case local hyperbolicity implies the existence of stable and unstable manifolds, which, when a perturbation is introduced, generically intersect in a transversal way, generating a homoclinic tangle, i.e. a chaotic region.

In addition to the maximum Lyapunov exponent, other $n-1$ positive or null Lyapunov exponents (n being the number of degrees of freedom) can be effectively computed, but using a more complicated recipe illustrated in Benettin *et al.* (1980). The number of positive Lyapunov exponents indicates the number of independent directions in phase space along which the orbit exhibits chaotic (or at least hyperbolic) behavior. A theorem by Pesin (1977) shows that the KS–entropy can be computed from the Lyapunov exponents $\mathcal{L}_i$ as:

$$h = \int \sum_{\mathcal{L}_i > 0} \mathcal{L}_i(\mathbf{x}) \mathrm{d}\mathbf{x} \ , \tag{5.4}$$

where the sum is over all positive Lyapunov exponents and the integral is over a specific region of phase space.

5.2.1 Numerical computation of the MLE

The numerical computation of the MLE is not straightforward, because one should integrate the system (5.1) for infinite time in order to compute the limit in (5.2). Moreover, if the system (5.1) is hyperbolic, $\delta\mathbf{p}$ and $\delta\mathbf{q}$ indefinitely grow with time, inevitably causing an overflow in the numerical computation of their evolution.

These problems have been brilliantly overcome by Benettin *et al.* (1980). Their suggested recipe is the following. Start with an arbitrary $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$; compute the evolution of $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$ up to some time T that is short

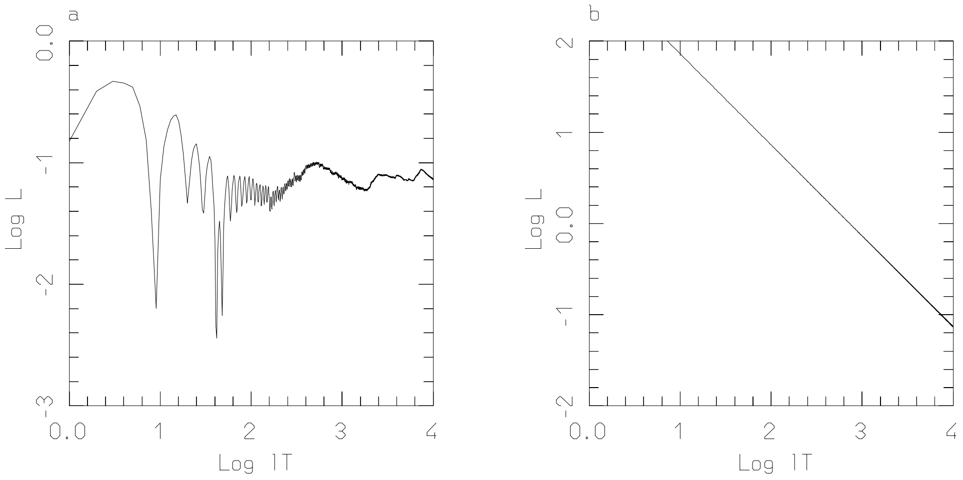


Figure 5.3: The evolution of the estimate of the Lyapunov exponent $\mathcal{L}(lT)$ for a chaotic (panel a) and for a regular orbit (panel b) of the standard map. Here $T = 1$ and l is the number of iterations of the standard map.

enough to avoid the overflow error on the computer; then, let $s_1 = \|\delta\mathbf{p}(T), \delta\mathbf{q}(T)\|/\|\delta\mathbf{p}(0), \delta\mathbf{q}(0)\|$ and define $\delta\mathbf{p}_1 = \delta\mathbf{p}(T)/s_1$ and $\delta\mathbf{q}_1 = \delta\mathbf{q}(T)/s_1$; finally use $\delta\mathbf{p}_1, \delta\mathbf{q}_1$ as new initial conditions $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$ and iterate the above procedure. In this way, one obtains a sequence of renormalization factors $s_1, s_2, \ldots, s_l$. It has been proven by Benettin *et al.* (1980) that

$$\mathcal{L} = \lim_{l\to+\infty} \frac{\sum_{j=1}^{l} \ln s_j}{lT} \ , \tag{5.5}$$

and that the result is independent of the choice of T.

To estimate the value of the above limit, one plots the quantity $\mathcal{L}(lT) \equiv (\sum_{j=1}^{l} \ln s_j)/lT$ as a function of lT on a logarithmic scale, and tries to guess the asymptotic behavior. Figure 5.3 gives an example of what one usually obtains for a chaotic orbit (panel a) and for a regular orbit (panel b). Despite the fact that the value of $\mathcal{L}(lT)$ is initially larger for the regular than for the chaotic orbit, in the regular case it seems to monotonically decrease with lT, while in the chaotic case it has large oscillations and eventually seems to stabilize around a definite value ($\sim 10^{-1.1}$ in Fig. 5.3). Then, it is reasonable to expect that in case (b) the limit for $lT \to +\infty$ of $\mathcal{L}(lT)$ is zero, while in case (a) it is $\sim 10^{-1.1}$.

In most cases, pictures like those of Fig. 5.3 are enough to discriminate between regular and chaotic orbits. However, one is never absolutely sure to guess the right limit. In some cases, the evolution of $\mathcal{L}(lT)$ initially decreases

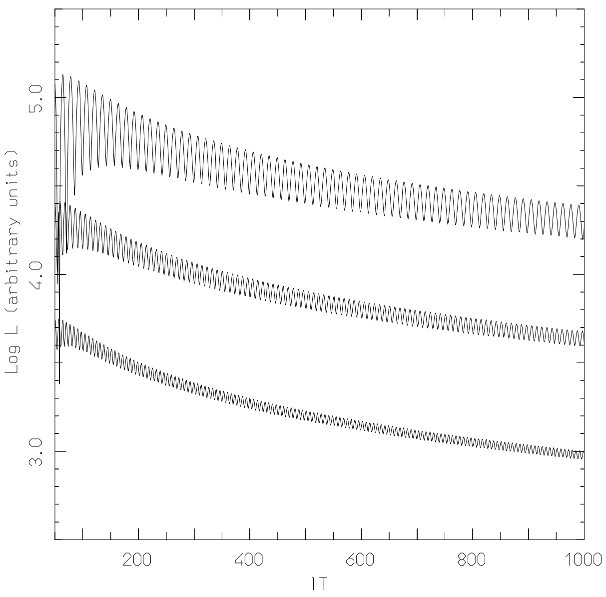


Figure 5.4: The time evolution of the estimate of the Lyapunov exponent $\mathcal{L}(lT)$ for three orbits with decreasing distances (from bottom to top) from the main resonance of the standard map. The size of the oscillations of $\mathcal{L}(lT)$ is an indicator of the distance from the most effective resonance.

as in Fig. 5.3b, but then asymptotically converges to some small, but positive value. Therefore, if the evolution of $\mathcal{L}(lT)$ is not followed for a long enough time, one can be erroneously led to guess that the Lyapunov exponent is zero. Actually, computing the evolution of $\mathcal{L}(lT)$ up to a finite value of lT, one can detect only the orbits with MLE larger than some quantity (related to the maximal value of lT used in the computation). Only by increasing the integration timespan can one detect orbits with smaller MLE. The detection of all chaotic orbits would require the exact computation of the Lyapunov exponent, i.e. infinite time.

The evolution of $\mathcal{L}(lT)$ is never perfectly monotonically decreasing: it always presents small oscillations. The amplitude of these oscillations is an indicator of the distance of the considered orbit from the most effective resonance. As an example, Fig. 5.4 shows the evolution of $\mathcal{L}(lT)$ for three orbits in the standard map; the linear scale for lT has been used to better see the periods of oscillations, and the value of Log$\mathcal{L}(lT)$ has been changed by a suitable constant in order to avoid superposition among the different curves. From bottom to top, the three curves correspond to orbits with decreasing distance

from the main resonance at $p_1 \sim 0$. The oscillations reveal that the vector $\delta\mathbf{p}, \delta\mathbf{q}$ in (5.1) is alternately contracted and stretched. This happens when the orbit passes close to the hyperbolic point of the resonance (at $p_1 = 0, q_1 = 0$ in the standard map 3.13), and the amount of stretching and contraction is evidently strongest for the orbit that passes closest to the hyperbolic point. Note also that the period of oscillation increases with decreasing distance from the resonance, which is a consequence of the fact that the circulation period of the resonant angle decreases. The oscillations of $\mathcal{L}(lT)$ as a function of lT might therefore help to guess whether an orbit will have at the end a positive or a zero MLE, without waiting until the final convergence of $\mathcal{L}(lT)$ to its asymptotic value. If the oscillations are almost undetectable, the orbit does not feel the close presence of a resonance, so that it very likely lies near an invariant torus, and its MLE is equal to zero; conversely, if large oscillations appear, the orbit is close to a resonance, so that it is probable (but not certain!) that the orbit is chaotic, with a positive MLE.

Also, if the evolution of $\mathcal{L}(lT)$ does not show significant oscillations, orbits with close initial conditions will have close values of $\mathcal{L}(lT)$, for any fixed lT. This is not true in the case of large oscillations of $\mathcal{L}(lT)$, because the period of these oscillations differs slightly from one orbit to another. As a consequence, the value $\mathcal{L}(lT)$ is a smooth function of the initial conditions only in the regions where resonances have negligible effects, while it is an irregular function in the regions that are chaotic or at least strongly affected by resonances. As an example, Fig. 5.5 shows the value of $\mathcal{L}(1000)$ for the standard map as a function of the initial action p_1, the initial value of q_1 being equal to 0 for all orbits. Following the discussion above, the figure can be interpreted as a map of the dynamical character of the system: the plateaux where $\mathcal{L}$ is almost constant (with $\mathcal{L}(1000) \sim 10^{-2.2}$) reveals the "regular" regions, where the relative volume of KAM tori is close to 1, while the peaks and the dips (and all discontinuities in general) mark the main chaotic or resonant regions.

To conclude this section, a final word of caution is necessary. Some authors, to avoid the computation of the linearized equations of motion (5.1), follow the Benettin *et al.* recipe (5.5) by computing $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$ as the difference between the evolutions of two trajectories of the system. This is a very dangerous way to proceed, as the result in this case depends on the time T at which the renormalization is done and on $\|\delta\mathbf{p}(0), \delta\mathbf{q}(0)\|$, unless both are extremely small and very high accuracy is used in all computations. Because of numerical instabilities, it is very easy to compute an incorrect value of the MLE (see Holman and Murray, 1996 for a discussion). It is therefore highly preferable to compute the evolution of $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$ using the linearized equations of motion in all the cases where the latter can be computed.

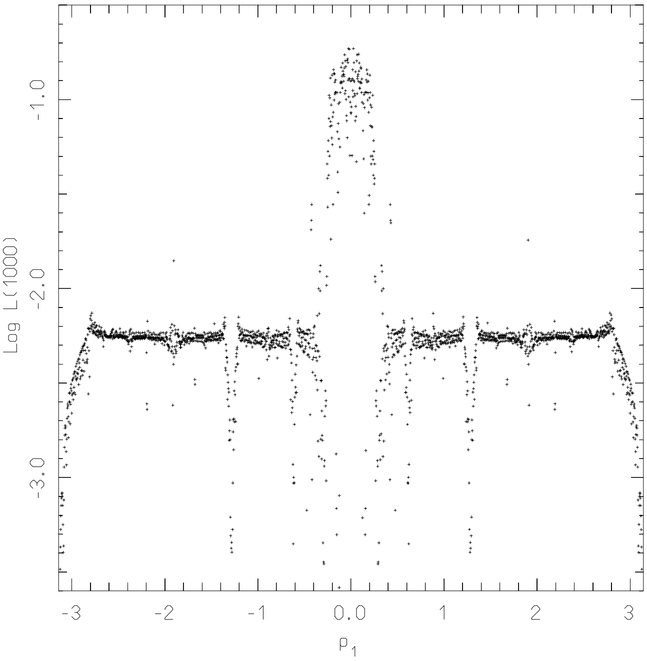


Figure 5.5: The value of the Lyapunov exponent, computed over 1000 iterations of the standard map, as a function of the initial value of the action. The parameter ε is here equal to 0.6, as in Fig. 3.3b.

5.3 Frequency analysis

This powerful tool for the detection of chaos has been introduced by Laskar (1990) to analyze the secular evolution of the planets (see Chapter 7), and it has been successfully used also for studying small body dynamics (Nesvorný and Ferraz-Mello, 1997b; Robutel and Laskar, 2001) and galactic dynamics (Papaphilippou and Laskar, 1998); its properties and applicability are detailed in Laskar *et al.* (1992a) and Laskar (1993).

The idea behind frequency analysis is simple. As seen in Chapter 3, orbits on KAM tori have constant frequencies; conversely chaotic orbits do not have well-defined frequencies, their actions and angles randomly moving inside the chaotic zone. The point is that the value of the frequencies can be numerically determined on a time interval, so that it is possible to check whether the frequencies change or stay constant from one time interval to another.

More precisely, given the evolution of an orbit $\mathbf{p}(t), \mathbf{q}(t)$ as resulting from a numerical integration, one numerically determines (following the procedure indicated in Section 5.3.1) the fundamental frequencies $\boldsymbol{\nu}(t_0)$ of the motion on

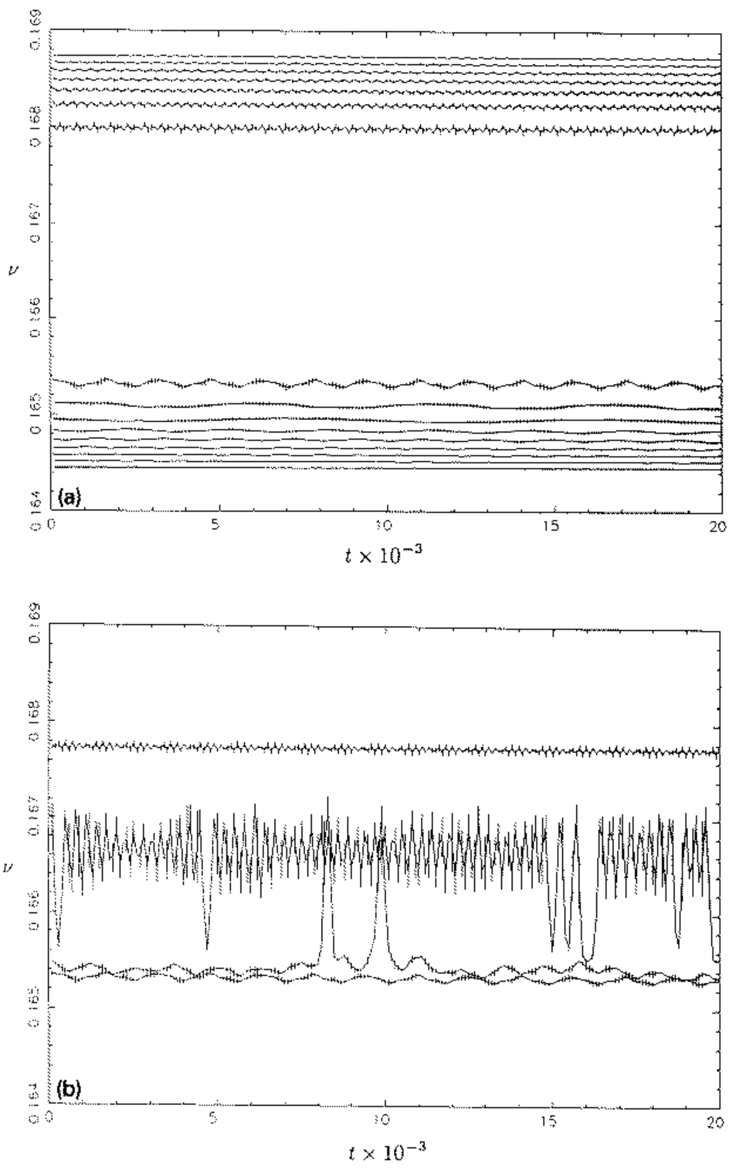


Figure 5.6: Time evolution of the frequency of orbits close to a 1/6 resonance in the standard map. The orbits for which the frequency is approximately constant with time lie on KAM tori, while the others are chaotic. Reprinted from Fig. 5 of Laskar *et al.* (1992a), with permission from Elsevier Science.

a time interval $[t_0, t_0+T]$. The duration T of the time interval must be chosen to be longer than the inverse of the slowest fundamental frequency, to achieve an accurate determination of the latter. Then one repeats the computation for different values of t_0, thus numerically determining the frequencies as a function of time. If the frequencies are constant, then the considered orbit is "regular", namely it either lies on a KAM torus or is a periodic orbit (Laskar, 1999). Conversely, if $\boldsymbol{\nu}(t_0)$ changes with t_0 there is a strong presumption that the orbit is chaotic. Figure 5.6 gives an example of this kind of analysis, for a sequence of orbits close to a 1/6 resonance in the standard map. The top panel shows the time evolution of the frequency of the angle q_1 for orbits which are on both sides and quite far from the resonance (located at $\nu = 0.1666\ldots$). For the most distant orbits from the resonance, straight lines are obtained, which implies that the time variations of the frequency are smaller than the resolution of the plot. Thus, these orbits presumably lie on KAM tori. Conversely, for the orbits that are closer to the resonance, the frequency ν shows time variations that are larger than the numerical accuracy by which the frequency is computed on each time interval (the latter being indicated by the vertical segments). These orbits therefore should not lie on KAM tori, but rather be weakly chaotic. For orbits even closer to the 1/6 resonance (panel b) the amplitude of frequency variations increases, suggesting a stronger chaos. The second orbit from the top is properly in the 1/6 resonance, and therefore its frequency has large irregular oscillations around the exact resonant value; note however a number of excursions of the frequency towards values smaller than 0.166, which correspond to temporary trapping of the orbit into the circulation region below the resonance. Conversely, the second orbit from the bottom spends most of the time below the resonance ($\nu \sim 0.1654$) but is twice temporarily trapped in the resonance (when $\nu \sim 0.1666$).

A second way to proceed to explore the dynamical structure of a system is to compute the frequencies as a function of the initial conditions. More precisely, one fixes the initial angles and constructs a fine grid of initial conditions $\mathbf{p}_0$ in the action space; then, for each initial condition one computes the corresponding frequencies $\boldsymbol{\nu}(\mathbf{p}_0)$ on the interval $[0, T]$, for some fixed large T. A theoretical result by Lazutkin (1973) shows that the frequencies on KAM tori can be fitted by a smooth ($\mathcal{C}^\infty$) function of the actions. Therefore, in the regions where KAM tori fill a large relative volume, the numerically determined function $\boldsymbol{\nu}(\mathbf{p}_0)$ must appear smooth. Conversely, such smoothness must be lost in the chaotic region, where the frequencies are not well defined and only rough approximations are computed on the $[0, T]$ interval. Figure 5.7 shows an example of this, again for the region around the 1/6 resonance in the standard map. The value of the frequency ν is plotted as a function of

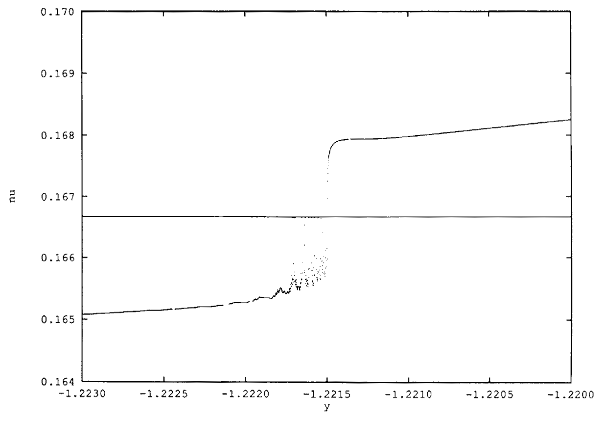


Figure 5.7: Frequency as a function of the initial value y of the action p_1, for orbits close to a 1/6 resonance in the standard map. The smooth parts of the graphe denote regular regions, while scattered points reveal chaotic regions. The horizontal line shows the exact resonant value of the frequency. Reprinted from Fig. 6 of Laskar *et al.* (1992a), with permission from Elsevier Science.

the initial value of p_1. The function $\nu(p_1)$ is smooth for $p_1 < -1.222$, slowly increasing with p_1; then it starts to show bumps and wiggles and subsequently becomes a collection of scattered points. The exact resonant value is crossed, and then $\nu(p_1)$ again becomes a smooth function of p_1.

Figures 5.6 and 5.7 show how effective is frequency analysis in detecting chaotic orbits and chaotic zones. With respect to the computation of Lyapunov exponents, frequency analysis requires in principle shorter integration times. Several periods of the slowest angle are enough to accurately determine the frequencies and understand their behavior with respect to time and/or initial conditions, while for the computation of the Lyapunov exponent one has to wait until the evolution of $\mathcal{L}(t)$ has reached its asymptotic limit. Moreover, from the frequency analysis it is straightforward to deduce which resonances are responsible for the origin of the detected chaos, while the computation of the Lyapunov exponent does not provide any indication in this sense. On the other hand, the numerical computation of the frequencies requires quite sophisticated software, in particular for the applications in Celestial Mechanics, where both slow and fast frequencies exist due to the degeneracy of the Keplerian problem. For these competing reasons, both Lyapunov exponent computations and frequency analysis are nowadays largely in use.

5.3.1 Numerical determination of the frequencies

This section is devoted to giving some indications of the practical implementation of frequency analysis. More details can be found in the original paper by Laskar *et al.* (1992a).

For an n-degree of freedom system, a regular orbit on a KAM torus is quasi-periodic with n independent frequencies $\nu_1, \ldots, \nu_n$. In principle, due to the coupling among the degrees of freedom, these n frequencies are present in the time evolution of each action and angle of the system. However, each frequency ν_j is in general the leading frequency of an angle q_j, so that, instead of looking for all independent frequencies in the time evolution of a unique variable, it is more convenient to look for the main frequency of each of the angles $q_j(t)$. If one looks for the frequency of libration inside a resonance, it is suitable to introduce the critical angle of the resonance φ_1 and its conjugate action I_1, as in Chapter 4, and then introduce the angle – say ϑ_1 – that defines in polar coordinates the position of I_1, φ_1 relative to the center of libration. The frequency of libration is the leading frequency of $\vartheta_1(t)$.

In order to avoid the problem raised by the definition of the angles modulo 2π, for each angle $q_j(t)$ the complex function $f_j(t) = \exp[\iota q_j(t)]$ is introduced. The leading frequency of $q_j(t)$ becomes of course the leading frequency of $f_j(t)$. Given a time interval $[t_0, t_0 + T]$ the frequencies of f_j cannot be determined with good accuracy using a simple fast Fourier transform, because $f_j(t)$ is generically not periodic on that interval. Instead, one computes the function

$$\Phi_j(\omega) \equiv \frac{1}{T} \int_{t_0}^{t_0+T} f_j(t) \exp(-\iota\omega t) \mathrm{d}t \tag{5.6}$$

and looks for the value ω^* for which $\Phi_j(\omega)$ assumes the maximal value. The value ω^* approximates the leading frequency of f_j. It is the *exact* frequency if the function $f_j(t)$ is simply periodic: in fact, for $f_j = \exp[\iota\nu t]$, (5.6) gives $\Phi_j(\omega) = \sin[(\nu - \omega)T/2]/[(\nu - \omega)T/2]$, which has an absolute maximum for $\omega = \nu$. If f_j has several independent frequencies $\nu_1, \ldots, \nu_n$, the value ω^* will not be exactly equal to the leading frequency; however, the "error" must be small if the frequencies are well distinct and the amplitude associated with the leading frequency is significantly larger than the amplitudes associated with the secondary frequencies.

In Celestial Mechanics applications, the problems are raised by the contemporary existence of fast and slow frequencies, associated respectively to the mean longitudes and to the perihelia and nodes of the bodies. From the point of view of numerical determination, this is equivalent to the case where the frequencies $\nu_1, \ldots, \nu_n$ are almost equal and have comparable associated amplitudes in the Fourier spectrum. In the original application by Laskar

(1990), this problem did not exist, because Laskar was considering only the secular evolution of the solar system, starting from the secular normal form (see Section 2.5), so that the fast angles were not present.

When one has to deal with the evolution of an orbit computed in the framework of the complete system (i.e. including also short-period terms), the best strategy is first to filter the evolution using a numerical procedure (Carpino *et al.*, 1987), in order to average over all short-period oscillations, and then to apply the frequency analysis on the filtered angles. The whole process is discussed in detail by Nesvorný and Ferraz-Mello (1997a) and efficiently allows the computation of the frequencies of the slow angles. However, in this way one cannot monitor the short-period frequencies.

5.4 Surrogates

A number of additional numerical tools for the detection of chaos have been recently introduced, as variations on the themes of Lyapunov exponents or of frequency analysis. Here we briefly review those that have been successfully used in Celestial Mechanics problems or seem promising for future applications.

5.4.1 Fast Lyapunov indicator

This was introduced by Froeschlé *et al.* (1997). For a given orbit of a system, one considers the linearized equations (5.1) and, starting from an arbitrary initial condition $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$, one computes the evolution $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$. The fast Lyapunov indicator (FLI) is the time T at which $\|\delta\mathbf{p}(t), \delta\mathbf{q}(t)\|$ is first equal to some large but arbitrarily fixed value R. Intuitively, the FLI is related to the Lyapunov exponent: orbits such that the system (5.1) is strongly hyperbolic have a short FLI; if (5.1) is weakly hyperbolic the FLI is longer, and if its motion of $\delta\mathbf{p}, \delta\mathbf{q}$ is bounded the FLI is infinite. However, the FLI evidently depends on the choice of the initial condition $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$, on the choice of R and on the choice of the canonical variables $\mathbf{p}, \mathbf{q}$ that are used to write the linearized equations. Therefore, in contrast to the MLE, it is not intrinsically related to the considered orbit and cannot be used as an absolute measure of its chaoticity.

However, once the choice of the canonical variables, the initial condition $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$ and the threshold value R are fixed, the computation of the FLI allows one to effectively compare the dynamical behaviors of different orbits. In this sense, the FLI can be considered as an indicator of the chaotic nature of an orbit relative to another one. To give absolute indications, it needs to

be calibrated using reference orbits, for which the chaotic nature has been determined, for instance computing the MLE.

The major quality of the FLI is the simplicity and the rapidity of the procedure required for its computation, which allows the investigation of a large set of orbits. The FLI technique has been successfully applied by Froeschlé *et al.* (1997) to investigate the dynamical evolution of all numbered asteroids, from which it has been possible to reveal the importance of three-body resonances for the origin of chaos in the asteroid belt (see Chapter 11).

5.4.2 Helicity and twist angles

These were introduced by Contopoulos and Voglis (1996). Given the linearized equations (5.1), one computes the orientation of the vector $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$ as a function of time. For an n-degree of freedom system, the orientation is defined by $n-1$ angles: the *helicity angles* $\Phi_1, \ldots, \Phi_{n-1}$. The mean values of the helicity angles with respect to time $\langle\Phi_1\rangle, \ldots, \langle\Phi_{n-1}\rangle$ in principle depend only on the orbit around which the linearized equations of motion (5.1) are computed, while they do not depend on the initial choice of the vector $\delta\mathbf{p}(0), \delta\mathbf{q}(0)$. Contopoulos and Voglis showed that for orbits in a uniform chaotic region the mean value of the helicity angles is invariant, while for regular orbits it smoothly changes with the initial conditions. This property allows one to easily distinguish regular from chaotic regions, once the mean value of the Helicity angles is computed on a regular grid of initial conditions. Of course the computation of the mean helicity angles requires in principle infinite computation time; however a quite precise indication of their values can already be effectively achieved on a quite short time interval, which makes this method appealing from the computational viewpoint.

In addition to the helicity angles, Contopoulos and Voglis considered also their time derivatives: the *twist angles*. Again, the mean values of the twist angles allow one to distinguish regular from chaotic orbits. They are invariant for orbits in a uniform chaotic zone, are identically equal to zero for orbits lying on KAM tori (for which all angles circulate), and are equal to the libration frequencies for orbits lying on resonant invariant tori (Chapter 4). Examples of the behavior of the mean value of the twist angle in the standard map can also be found in Froeschlé and Lega (1998).

5.4.3 Mean exponential growth factor of nearby orbits

The formula (5.2) for the computation of the maximum Lyapunov exponent can be rewritten in integral form as:

$$\mathcal{L} = \lim_{t\to+\infty} \frac{1}{t} \int_0^t \frac{\dot{\delta}(t')}{\delta(t')} \mathrm{d}t' \ , \tag{5.7}$$

where $\delta = \|\delta\mathbf{p}, \delta\mathbf{q}\|$ and $\delta\mathbf{p}(t), \delta\mathbf{q}(t)$ are solutions of (5.1); $\dot{\delta}$ denotes, as usual, the time derivative of δ. The mean exponential growth factor of nearby orbits (MEGNO) was introduced by Cincotta and Simó (2000) as the quantity

$$Y(t) = \frac{2}{t} \int_0^t \frac{\dot{\delta}(t')}{\delta(t')} t' \mathrm{d}t' \ . \tag{5.8}$$

The evolution of this quantity as a function of time allows the determination of the dynamical character of the considered trajectory (the trajectory $\mathbf{p}(t), \mathbf{q}(t)$ around which the variational equations (5.1) are computed). In fact, for a quasi-periodic trajectory, for which δ grows linearly with time, $Y(t)$ oscillates with bounded amplitude around the value 2. Conversely, for a chaotic trajectory, for which δ diverges exponentially with time and σ is the maximum Lyapunov exponent, $Y(t)$ oscillates around the linearly divergent solution $y = \sigma t$. Thus, introducing the running time-average of the MEGNO:

$$\bar{Y}(t) = \frac{1}{t} \int_0^t Y(t') \mathrm{d}t' \ , \tag{5.9}$$

one has

$$\lim_{t\to+\infty} \bar{Y}(t) = 2 \tag{5.10}$$

for a quasi-periodic trajectory and

$$\lim_{t\to+\infty} \bar{Y}(t) \sim \frac{\sigma}{2} t \tag{5.11}$$

for a chaotic trajectory.

At first sight, the time-average of the MEGNO seems to be perfectly equivalent to the maximum Lyapunov exponent. Indeed, it shares the same nice property of depending solely on the considered trajectory, and not on the coordinates and on the metric used for the computations. However, Cincotta and Simó showed that $\bar{Y}(t)$ converges to its limit value faster than the estimate of the Lyapunov exponent. This is because the latter keeps a long memory of the initial transient evolution of the tangent vector $\delta\mathbf{p}, \delta\mathbf{q}$; on the contrary, the integral (5.8) gives a preferential weight to the late evolution of $\delta\mathbf{p}, \delta\mathbf{q}$, thus losing more rapidly the memory of the initial transient evolution.

Cincotta and Giordano (2001) have provided numerous examples of the application of the MEGNO on systems with two and three degrees of freedom. In all cases, the MEGNO was found to be advantageous with respect to the computation of the maximum Lyapunov exponent: in equal computing time, the MEGNO allowed the detection of finer chaotic regions than the MLE.

5.4.4 Mean, maximal and minimal values of the actions

The idea of this approach is inspired by the frequency analysis method. If an orbit has a regular evolution (it is periodic or lies on a KAM torus), not only are the frequencies constant, but also the behavior of the actions repeats periodically or quasi-periodically (see for instance Fig. 5.1a). Therefore, the minimal, the maximal and the mean values of the actions on a time interval $[t_0, t_0 + T]$ do not depend on the initial time t_0 of the interval, provided T is large enough.

From this consideration, an easy criterion to distinguish regular from chaotic orbits is derived: regular orbits are those for which the mean, maximal or minimal values of the actions do not change from one time interval to another; chaotic orbits are those for which these values change with t_0. This criterion was first used by Laskar (1994) to illustrate the long-term evolution of the planets, and since then has been repeatedly applied (see for instance Morbidelli, 1997, Morbidelli and Nesvorný, 1999) to point out slow changes of the orbital elements of asteroids and Kuiper belt objects. The advantage of this technique with respect to frequency analysis is its much easier implementation; moreover it gives more easily readable information on the long-term changes of the mean orbital elements (semimajor axis, eccentricity and inclination). Conversely, frequency analysis allows easy identification of the resonance(s) that are responsible for the origin of chaos, which cannot be obtained by simply monitoring the actions.

The maximal, minimal and mean values assumed by the actions on a time interval $[0, T]$ can also be studied as a function of the initial conditions, similarly to what can be done for the frequencies. This again allows one to distinguish regular from chaotic regions. An extensive discussion, and examples based on the standard map, can be found in Contopoulos *et al.* (1997).

Chapter 6

INTERACTIONS AMONG RESONANCES

6.1 Two degrees of freedom

In Chapter 4 we have seen that a resonance located in $\mathbf{p}_0$ in reality influences the dynamics in a neighborhood centered on $\mathbf{p}_0$, whose size is of order $\sim \sqrt{c}$, where c is the coefficient of the resonant harmonic. As a consequence, resonances located at different places in the action space are effectively separated only if the coefficients of the corresponding resonant harmonics are small enough, otherwise their domains of influence overlap.

To clarify, consider the quasi-integrable Hamiltonian

$$\mathcal{H}(p_1, p_2, q_1, q_2) = \frac{p_1^2}{2} + 2\pi p_2 + \varepsilon[\cos(q_1) + \cos(q_1 - q_2)] \tag{6.1}$$

and consider the resonances $\dot{q}_1 = 0$ and $\dot{q}_1 - \dot{q}_2 = 0$, located respectively at $p_1 = 0$ and $p_1 = 2\pi$ for each value of p_2. To study the dynamics of this system, to a first approximation we can consider each resonance separately, taking into account only its corresponding harmonic and neglecting the other harmonic. More precisely, we use the integrable single resonance models $\mathcal{H} = p_1^2/2 + 2\pi p_2 + \varepsilon\cos(q_1)$ and $\mathcal{H} = p_1^2/2 + 2\pi p_2 + \varepsilon\cos(q_1 - q_2)$ to compute the Poincaré section around $p_1 = 0$ and $p_1 = 2\pi$ respectively. The resulting description of the dynamics is illustrated in Fig. 6.1. If ε is small, the separatrices of the two resonances are far; the domains, delimited by the dashed lines, where the dynamics is significantly influenced by each resonance, are very far apart. The assumption that the resonances are separated turns out to be a good approximation; from Chapter 4, we know that in reality the separatrix of each resonance is split, giving rise to a small chaotic region, but we expect that

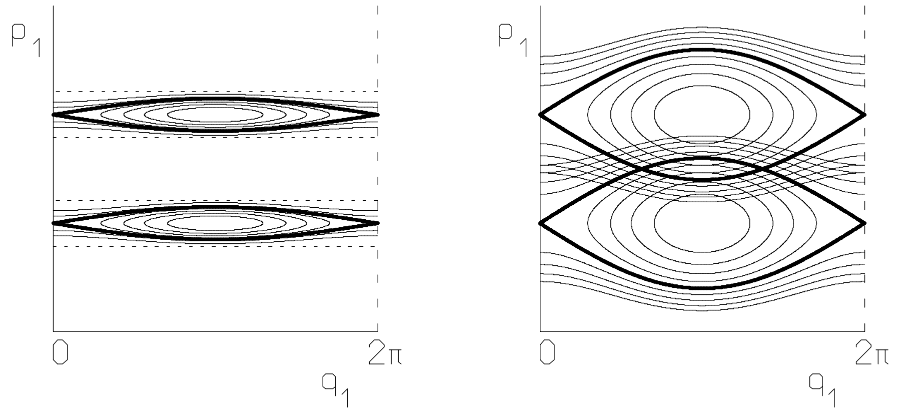


Figure 6.1: Poincaré sections of two resonances, computed using integrable single resonance models. When their sizes are small, the resonances are well separated, and the single resonance approximations are reliable (panel a: left). Conversely, when their sizes are large, resonances overlap (panel b: right). In the latter case, the generation of large-scale chaos may be expected.

many KAM tori exist between the two resonances, so that each chaotic region is locally confined. In particular, it is not possible for the motion to pass from one resonance to the other.

The situation is completely different when ε is large (Fig. 6.1b). In this case the domains where the dynamics is strongly influenced by either resonance overlap each other, and the separatrices of the two resonances, as computed on the Poincaré section with the corresponding integrable models, appear to intersect. The approximation that each resonance is isolated therefore breaks: the global dynamics cannot be studied using the integrable single resonance models. It is reasonable to expect that in this situation KAM tori do not exist between the two resonances, so that the chaotic regions associated to the separatrices of the two resonances are in reality connected, and approximately extend over all the space occupied by the two resonances. Roughly speaking, an initial condition in the overlapping region does not know which resonance it belongs to, and "hesitates" about which guiding trajectory (the curves in Fig. 6.1b) it should follow. As a result, the evolution of an initial condition in the chaotic region can freely pass from one resonance to the other one, the action p_1 evolving from 0 to 2π and vice versa and the two resonant angles q_1 and $q_1 - q_2$ alternatively librating and circulating.

The intersection of the separatrices of different resonances, each computed

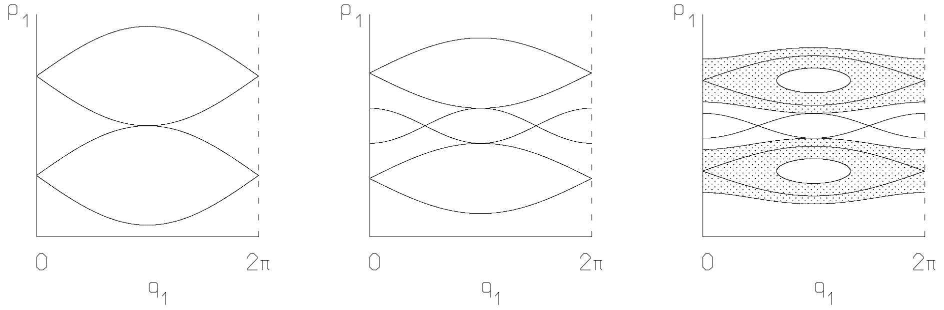


Figure 6.2: Left: the simple overlapping criterion between the two main resonances at $p_1 = 0$ and $p_1 = 2\pi$ ($\varepsilon = 2.47$). Middle: overlapping criterion taking into account also the resonance at $p_1 = \pi$ ($\varepsilon = 1.46$). Right: overlapping criterion taking into account the width of the chaotic layers (dotted regions) of the two main resonances ($\varepsilon = 1.2$).

using integrable single resonance models, has been considered by Chirikov (1960, 1979) and Contopoulos (1966) as a practical criterion to quantitatively estimate the threshold value of ε that corresponds to the disappearance of KAM tori and the transition to global chaos. This criterion is now usually known as the *Chirikov criterion*, and its accuracy has been tested by Chirikov (1979) on the standard map (3.13). Taking into account only the two main resonances $\dot{q}_1 = 0$ and $\dot{q}_1 - \dot{q}_2 = 0$ at $p_1 = 0$ and $p_1 = 2\pi$, Chirikov found a critical value of $\varepsilon \sim 2.47$. The real value for the transition to global chaos, which corresponds to the destruction of the *golden KAM torus* (the torus with frequency ratio $\dot{q}_1/\dot{q}_2 = (\sqrt{5}-1)/2$), is $\varepsilon \sim 0.9716$, as determined numerically by Green (1979). Therefore the result provided by the Chirikov criterion seems only partially satisfactory and very approximate from the quantitative viewpoint. In general the Chirikov criterion underestimates the threshold of transition to global chaos, in the sense that the actual transition occurs for a value of the perturbing parameter which is smaller than the one predicted by the criterion. This is due mainly to two facts: first, one neglects all the resonances whose harmonics are not present in the original Hamiltonian but are generated after a suitable number of perturbation steps (see Chapter 2), with coefficients of higher order in ε, and, second, one completely neglects the mutual interactions and deformations induced among the resonances. Chirikov tried to improve from the quantitative point of view the results of his criterion by taking into account more and more perturbation terms. Taking into account the resonance $2\dot{q}_1 - \dot{q}_2 = 0$ at $p_1 = \pi$, whose harmonic coefficient is of order ε^2, the estimated threshold decreases to $\varepsilon = 1.46$, and taking into account also the

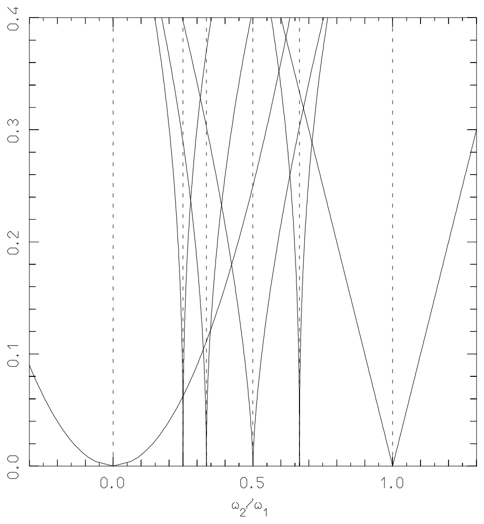


Figure 6.3: Sketch of the overlapping of resonances as a function of the perturbation parameter for a two-degree of freedom system. The resonance widths depend on the size ε of the perturbation. If the perturbation is larger than a given threshold, resonance overlapping occurs. For ε sufficiently small, resonances cannot overlap, and invariant tori fill almost the entire volume.

width of the chaotic layers around the two main resonances one gets $\varepsilon = 1.2$. These attempts are summarized in Fig. 6.2 and indicate that the accuracy of the criterion is improved when more and more resonant perturbation terms are considered.

From the discussion above, we can sketch the global dynamical behavior of a two-degree of freedom system as a function of the perturbation parameter ε as done in Fig. 6.3. The figure shows on the x axis the ratio ω_1/ω_2 of the frequencies of q_1 and q_2 and on the y axis the size of the perturbation. The resonances are located at rational values of the frequency ratio, and therefore are represented by vertical dashed lines in the plot. The solid curves show the amplitude of each considered resonance, measured in terms of the frequency ratio. The latter is obtained as follows: in the integrable single resonance model one computes the values $\mathbf{p}^+$ and $\mathbf{p}^-$ of the action at the apex of the upper and lower separatrix (for instance $p_1 = \pm 2\sqrt{\varepsilon}$ for the $\dot{q}_1 = 0$ resonance in the Hamiltonian (6.1)), and then shows the values of the unperturbed frequency ratio in $\mathbf{p}^+$ and $\mathbf{p}^-$. Because the size of the coefficients of the resonant harmonics increase with the size of the perturbation and the widths of the resonances scale as the square roots of such coefficients, the resonance amplitudes in Fig. 6.3 have a typical V shape. As a consequence, if the size of

the perturbation is smaller than a threshold value (the latter being in general a function of the frequency ratio), the resonances do not overlap, and KAM tori are expected to exist among the resonances. The orbits cannot pass from one resonance to another and are therefore confined in the action space. Conversely, above the threshold, resonances overlap and KAM tori disappear. The motion can therefore pass from one resonance to the other, and the frequency ratio evolves with time following a sort of random walk: the width of the steps of the walk is approximately given by the resonance amplitudes, and the timescale for one step is of the order of the typical period of libration of the resonant critical angle. In this book, we will refer to this phenomenon as *Chirikov diffusion*. By the word *diffusion*, we do not imply that the evolution of the frequency ratio strictly follows a diffusion equation, but simply that it has macroscopic variations, revealing transport in the action space.

It should be noted that Fig. 6.3 has been drawn only for resonances up to some finite order, while in principle resonances of every order should be considered. However, as one can see from the figure, the widths of the resonances rapidly decrease with the resonance order. This comes from the fact that the coefficients of the resonant harmonics decrease exponentially with the order of the considered resonances, while the mutual distance among resonances of order K decreases as $1/K$. As a consequence, high-order resonances play a negligible role for what concerns global resonance overlapping. The resonances that may overlap are of order smaller than some threshold value. This is a general fact also for systems with more than two degrees of freedom, so we will come back to this important point in Section 6.3.

6.1.1 Heteroclinic intersections

As we have seen above, the Chirikov criterion for the transition to global chaos is very attractive and intuitive; unfortunately it is a very heuristic argument and a complete rigorous theory does not yet exist. The idea of resonance overlapping can be founded on the concept of *heteroclinic intersection*. This is a very precise mathematical concept: it is the intersection of the stable and the unstable manifolds of two different resonances (contrary to the homoclinic intersection discussed in Chapter 4, which is between the stable and the unstable manifolds of the same resonance). Unfortunately, it is much more difficult to study heteroclinic intersections than homoclinic intersections. Homoclinic intersections can be studied in the framework of perturbation of the integrable case, where the stable and the unstable manifolds coincide. Conversely, heteroclinic intersections between different resonances cannot be studied using perturbation theory, because an integrable approximation which takes into account both resonances does not exist, at least in the generic case. The

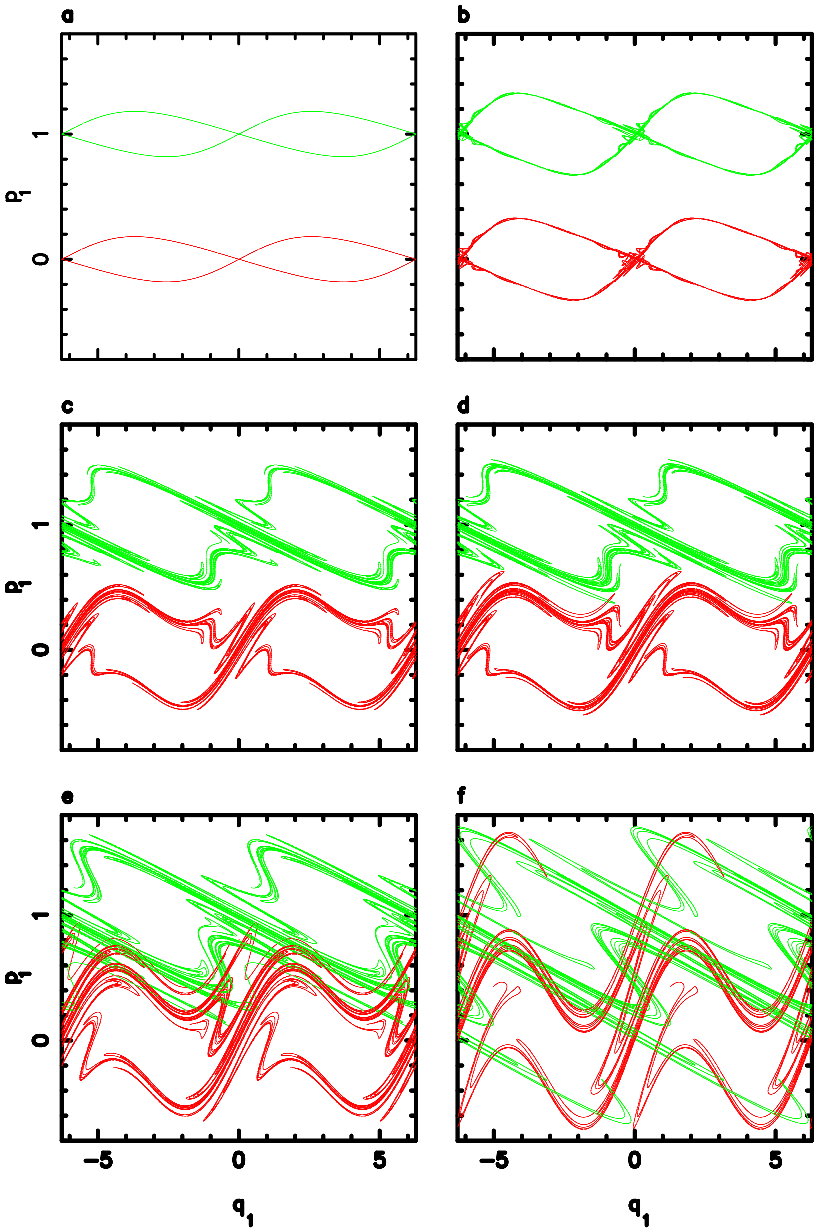


Figure 6.4: Homoclinic and heteroclinic intersections between the two main resonances of the standard map, for several values of the parameter ε. See text for comments.

portrait is indeed very complicated; up to now, only numerical computations can give us a clear view of what happens. Remember in fact from Chapter 4 that one can numerically compute the stable and the unstable manifolds of hyperbolic periodic orbits. Figure 6.4 shows how heteroclinic intersections occur between the manifolds of the two main resonances of the standard map. Panel (a) is done for $\varepsilon = 0.314$ for which the dynamics is quasi-integrable: the stable and the unstable manifold of each of the two resonances seem to coincide at the resolution of the picture. Panel (b) is for $\varepsilon = 0.942$. The resonances are still well separated; the homoclinic tangle of each resonance is now visible near the hyperbolic equilibrium points of the two resonances. On panel (c) ($\varepsilon = 1.632$) the lobes of the stable and the unstable manifolds of the two resonances are now very close, but still don't intersect, at least for what can be seen on the figure (the manifolds have infinite length, but evidently only a finite part of them can be computed and plotted). In this and in the following panels only the unstable manifold of the resonance at $p_1 \sim 0$, and the stable one of the resonance at $p_1 \sim 2\pi$ are plotted, otherwise the portrait would be too confused. On panel (d) ($\varepsilon = 1.696$) the lobes finally intersect transversally in what is called the *heteroclinic point.* One can apply the same reasoning done in Chapter 4 for the homoclinic points, in order to show the existence of infinite heteroclinic points and of a heteroclinic tangle. Unfortunately, on the pieces of manifold computed here, the multiple heteroclinic intersections and the heteroclinic tangle cannot yet be seen. They are more and more evident on the successive images (panels (e) and (f)) obtained for $\varepsilon = 2.198$ and $\varepsilon = 3.14$.

It is interesting to compare the threshold value found here for heteroclinic intersections, i.e. $\varepsilon_H \sim 1.66$, with the real threshold value for the transition to global chaos, i.e. $\varepsilon_G = 0.9716$ (the threshold for the disruption of the golden torus), and with the critical value provided by the Chirikov criterion, i.e. $\varepsilon_C = 2.47$. Notice that $\varepsilon_H \gg \varepsilon_G$ because we take into account only the two main resonances, neglecting all the higher-order resonances in between; if we had looked for heteroclinic intersections of the complete resonant chain, we would have found a critical value much closer to the real one (see Olvera and Simo, 1987). Alternatively, one should compute the stable and the unstable manifolds of the two main resonances on their infinite length, which is evidently impossible. Conversely, $\varepsilon_H \ll \varepsilon_C$ since the Chirikov criterion takes into account only the unperturbed resonance amplitude, neglecting mutual perturbations and the width of the homoclinic lobes. The computations illustrated in Fig. 6.4 show how complicated is the real portrait of the dynamics and one can imagine how difficult it would be to describe these phenomena in a rigorous mathematical way; it is therefore not surprising that a complete the-

oretical basis of the Chirikov criterion and of the Chirikov model of diffusion has not been provided up to now.

6.2 More than two degrees of freedom

Understanding the global dynamics of systems with more than two degrees of freedom is much more difficult than in the two-degree of freedom case. The *geography of resonances*, namely the relative location of the resonances in the frequency space, is in fact much more complicated. In the three-degree of freedom case, for instance, resonance locations can be represented on the plane with coordinates ω_1/ω_2 and ω_1/ω_3, where $\omega_1, \omega_2, \omega_3$ are the frequencies of the system. All possible lines with rational slope on such a plane represent resonances of multiplicity 1 (see Section 1.10 for a definition of multiplicity), and form what is usually called the *resonance web* or *Arnold web*. It is then evident that, differently from the two-degree of freedom case, resonant lines can intersect, the intersection points corresponding to resonances of multiplicity 2. Because of these intersections, resonance overlapping always occurs (at least locally in the vicinity of resonances of multiplicity larger than 1) so that the extension of the Chirikov criterion is not straightforward. Moreover, we know from Chapter 3 that the existence of KAM tori does not prevent transport in action space. As a consequence, while in two degrees of freedom only two situations occur – globally stable, where resonant motion is confined by KAM tori; and globally unstable, caused by resonance overlapping – in more than two degrees of freedom the dynamical behavior is more complicated.

Nevertheless, for small enough perturbations, Nekhoroshev (1977, 1979) has achieved a rigorous result of practical global stability of multidimensional systems.

6.2.1 The Nekhoroshev theorem

The Nekhoroshev theorem concerns quasi-integrable, n-degree of freedom Hamiltonians of type $\mathcal{H}(\mathbf{p}, \mathbf{q}) = \mathcal{H}_0(\mathbf{p}) + \varepsilon\mathcal{H}_1(\mathbf{p}, \mathbf{q})$, where $\mathbf{p}, \mathbf{q}$ are action–angle variables defined in a domain $\mathcal{D} \equiv \mathcal{G} \times \mathbf{T}^n$; $\mathcal{G}$ (space of actions) is a domain of $\mathbf{R}^n$ and $\mathbf{T}^n$ (space of angles) is the n-dimensional torus. In the following we denote by $\mathcal{G} - \Delta$ the set of points $\mathbf{p}$ which are contained in $\mathcal{G}$ together with a neighborhood of radius Δ. With these premises, the Nekhoroshev theorem can be stated in the following form:

Let $\mathcal{H}(\mathbf{p}, \mathbf{q}) = \mathcal{H}_0(\mathbf{p}) + \varepsilon\mathcal{H}_1(\mathbf{p}, \mathbf{q})$ *be analytic in* $\mathcal{D} \equiv \mathcal{G} \times \mathbf{T}^n$*, with* $\mathcal{G} \subset \mathbf{R}^n$ *open and bounded. Consider the matrix* $C(\mathbf{p})$ *defined by* $C_{ij}(\mathbf{p}) = \frac{\partial^2 \mathcal{H}_0}{\partial p_i \partial p_j}(\mathbf{p})$,

and assume the convexity hypothesis:

$$C(\mathbf{p})\mathbf{v}\cdot\mathbf{v} \neq 0 \quad \forall \mathbf{p} \in \mathcal{G} \text{ and } \forall \mathbf{v} \in \mathbf{R}^n \setminus \mathbf{0} \ . \tag{6.2}$$

Then, there exist positive constants ε_*, α, β, a and b such that for any $\varepsilon < \varepsilon_*$ one has

$$\|\mathbf{p}(t) - \mathbf{p}(0)\| \leq \Delta \equiv \alpha\varepsilon^a \ ,$$

for all $\mathbf{p}(0) \in \mathcal{G} - \Delta$ and for all $|t| \leq T(\varepsilon)$, where

$$T(\varepsilon) = \beta \left(\frac{\varepsilon_*}{\varepsilon}\right)^{\frac{1}{2}} \exp\left(\frac{\varepsilon_*}{\varepsilon}\right)^b \ . \tag{6.3}$$

A few comments are in order. The Nekhoroshev theorem does not exclude the possibility of chaotic motions. Indeed, the actions $\mathbf{p}$ can possibly change in a chaotic way: the theorem just states that these changes are bounded by a quantity Δ up to time T. Slow drifts can force the actions to change more than Δ with respect to the initial conditions only after a time larger than T, as sketched in Fig. 6.5. The important point is that the stability time T grows exponentially with respect to $\varepsilon_*/\varepsilon$. Therefore, as soon as ε is somewhat smaller than the threshold ε_*, the stability time becomes extremely long and can possibly exceed the physical lifetime of the system (for instance the age of the Solar System), thus providing a result of *practical* stability. We stress that such an important stability result is achieved *uniformly* for every initial condition $\mathbf{p}, \mathbf{q}$ with $\mathbf{p}$ in $\mathcal{G} - \Delta$. Obviously, orbits closer than Δ to the borders of the action domain are excluded, since they can escape from $\mathcal{G}$ in a short time. Concerning the hypotheses of the theorem, we remark that it is crucial to deal with analytic Hamiltonians. For reasons which will be explained below, one should expect that differentiable Hamiltonians have a stability time T which is a power of $\varepsilon_*/\varepsilon$, instead of having an exponential dependence. The role of the convexity condition (6.2) will also be explained below; it is worth mentioning that this condition can be weakened, using a more technical condition called *steepness* (Nekhoroshev, 1979), the discussion of which goes beyond the scope of this book.

The Nekhoroshev theorem is much more than the celebrated long time stability result reported in the statement above: it proves the existence of a specific structure of the phase space from which the stability result is derived. However this fact is usually hidden in the proof of the theorem, thus receiving little attention. In order to bring this structure into light, we sketch in the following the basic construction of the Nekhoroshev theorem.

As we have seen in Chapter 2, the harmonics $\varepsilon c_k(\mathbf{p})\exp(\imath\mathbf{k}\cdot\mathbf{q})$ in the perturbation $\varepsilon\mathcal{H}_1$ can be reduced to higher order in ε by means of the Lie

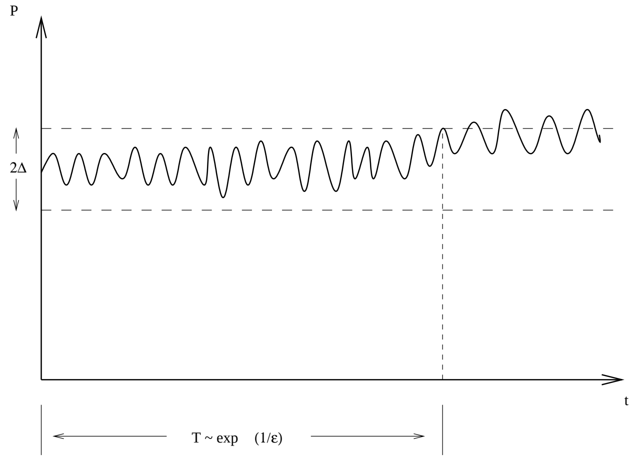


Figure 6.5: The Nekhoroshev theorem: whatever the initial conditions, the actions may evolve chaotically, but the time T required to change by more than a quantity $\Delta(\varepsilon)$ is exponentially long in $1/\varepsilon$. If the perturbation parameter ε is small, the Nekhoroshev theorem therefore gives a result of "practical" stability, the time T being longer than the physical lifetime of the system under consideration. Reprinted from Fig. 1 of Morbidelli and Guzzo (1996), with permission from Kluwer Academic Publishers.

transforms only away from the corresponding resonances $\mathbf{k} \cdot \boldsymbol{\omega} = 0$. Near the resonance the harmonic can't be removed and must be kept in the resonant normal form. Since resonances are dense in the action space, in any given open subset of $\mathcal{G}$ an infinite number of harmonics cannot be removed from the perturbation. This fact prevents in general the integrability of the system (Poincaré, 1892). The idea of Nekhoroshev is to consider resonances only up to a threshold order K.

This approach is very important. Indeed, on one hand, the number of resonances up to a given order is finite, and so each open subset of the space of actions contains only a finite number of resonant lines, as in Fig. 6.6. On the other hand the analyticity of the Hamiltonian implies that the neglected terms, corresponding to resonances of order larger than K, have a size not exceeding $\exp(-K\sigma)$ (for some positive σ). It will be shown that the threshold K can be chosen as large as $1/\varepsilon^b$ (with some positive $b < 1$), so that the neglected terms turn out to be exponentially small in $1/\varepsilon^b$. This will be a key point to obtain a stability time $T(\varepsilon)$ that depends exponentially on $1/\varepsilon^b$, as in (6.3). Note that if the Hamiltonian is only r times differentiable (instead of analytic), one expects that some of the neglected terms are of size K^{-r}, i.e. $1/\varepsilon^{rb}$. As a

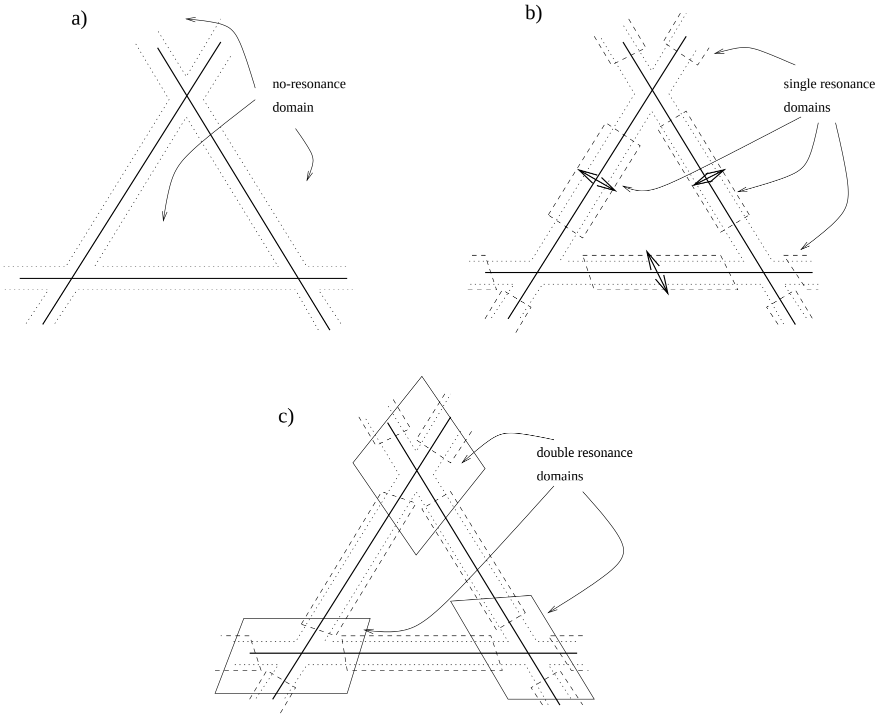


Figure 6.6: Sketch of the geometric construction of the Nekhoroshev theorem for a three-degree of freedom system in the frequency plane of coordinates ω_1/ω_3 and ω_2/ω_3. The three bold lines denote the locations of three resonances of multiplicity 1 and form a mesh of the Arnold web. The points where these resonances cross each other correspond to the locations of resonances of multiplicity 2. See text for discussion. Reprinted from Fig. 2 of Morbidelli and Guzzo (1996), with permission from Kluwer Academic Publishers.

consequence, one would obtain a stability time $T(\varepsilon)$ that is not larger than $\sim 1/\varepsilon^{rb}$.

For the sake of simplicity, we now restrict our attention to the three-degree of freedom case. Moreover, for a given value of the Hamiltonian, we represent the resonant structure of the system on the frequency plane of coordinates ω_1/ω_3 and ω_2/ω_3, where $\boldsymbol{\omega} = \mathrm{grad}_{\mathbf{p}}\mathcal{H}_0$. These choices allow us to draw in Fig. 6.6 nice explanatory pictures. Remember that in the frequency plane the resonances are represented by straight lines with rational slopes.

We first define a *nonresonant domain*, as the set of frequencies which are far enough from all resonances up to order K. In Fig. 6.6a it is the discon-

nected domain, bounded by the dotted lines. More precisely, it is defined as the set of frequencies $\boldsymbol{\omega}$ such that $|\mathbf{k}\cdot\boldsymbol{\omega}| > \sqrt{\varepsilon}$ for all $\mathbf{k}$ with $|\mathbf{k}| \leq K$.[1] In the nonresonant domain, one can construct the Birkhoff normal form, eliminating all harmonic terms in the perturbation $\varepsilon\mathcal{H}_1$ of type $e^{\iota\mathbf{k}\cdot\mathbf{q}}$ with $|\mathbf{k}| \leq K$. Indeed, by construction, all these terms are nonresonant in such a domain. The resulting Hamiltonian depends only on the new actions, apart from a remainder $\mathcal{R}_K$, made (essentially) of harmonic terms of order larger than K, which are exponentially small. Therefore, neglecting at most exponentially slow diffusion forced by $\mathcal{R}_K$, one can conclude that the frequencies of the system don't change with time in the nonresonant domain.

As a second step, we consider the *single resonance domains*, bounded by the dashed lines in Fig. 6.6b, which are characterized by the presence of only one resonance of order smaller than K. One can then construct a single resonance normal form, eliminating all nonresonant terms of order smaller than K. Then one reduces the Hamiltonian to having only one resonant term (of order $|k| \leq K$) and a remainder $\mathcal{R}_K$ made of terms of order larger than K, which is, again, exponentially small. Neglecting the remainder, such a Hamiltonian is still integrable but depends on one resonant angle, so that the actions (and, by consequence, the frequencies) are no longer fixed. They change, driven by the resonant harmonic, along what is usually called the *fast drift direction*, sketched in Fig. 6.6b by an arrow. The convexity hypothesis (6.2) guarantees that the resonant motion is bounded, as in the case of the pendulum. In fact, for a convex Hamiltonian, one can prove that the fast drift direction is transversal to the resonant line. Therefore, following indefinitely the fast drift direction, the motion would enter the nonresonant domain. But this is impossible, because in the nonresonant domain the frequencies are fixed, as explained above. On the other hand, transversal motion with respect to the fast drift direction can be forced only by the nonintegrable remainder $\mathcal{R}_K$, so that it is exponentially slow. This slow motion is usually called *Arnold diffusion.* Also the separatrix splitting is caused by the remainder $\mathcal{R}_K$, so that its amplitude is at most exponentially small (Neishtadt, 1984).

As a third step, we consider the *double resonance domains*, centered around the resonance crossings (the resonances of multiplicity 2) bounded by the light continuous lines in Fig. 6.6c. In such domains, the resonant normal form has two independent resonant terms of order smaller than K. Then, the resonant normal form is generically nonintegrable, and one expects that these domains

[1] Recall from Chapter 4 that the size of a resonance is proportional to the square root of the coefficient of the resonant harmonic (ε in this case); therefore, in the definition of the nonresonant domain, one must exclude a strip of width $\sqrt{\varepsilon}$ around the considered resonant lines.

are characterized by strongly chaotic motions and that frequencies can move in any direction of the plane around the resonance crossing. However, again this chaotic motion is bounded. Indeed, if frequencies moved far enough from the double resonance point, they would enter either the nonresonant domain or one of the single resonance domains. But this is impossible, since in the nonresonant domain frequencies are fixed, and in the single resonance domains frequencies can change only along the fast drift direction.

In conclusion, neglecting the exponentially small remainder $\mathcal{R}_K$, for each initial condition, the motion is confined within one of the resonance domains. As a consequence, frequencies (and actions) can change by at most a quantity equal to the radius of the double resonance domains (in n degrees of freedom the worst case corresponds to the domains constructed around resonances of multiplicity $n-1$). It can be proven that this radius is proportional to ε^a for some positive $a < 1$, (a decreasing with increasing number of degrees of freedom).

Moreover, in order to have a consistent picture as in Fig. 6.6, the number of resonances of order smaller than K must not be too large, otherwise there wouldn't be a place for the nonresonant domain, and the construction of Fig. 6.6 would be impossible. The fact that the largest resonance domains are of order ε^a gives an upper bound of type $1/\varepsilon^b$ on the choice of K, as anticipated above.

Finally, we take into account the exponentially small remainder $\mathcal{R}_K$. It is evident that this remainder can force diffusion in every direction of frequency space, but only with exponentially small speed. Then the result concerning bounded motion, deduced neglecting $\mathcal{R}_K$, will be true in principle only up to exponentially long times.

The scheme for the proof of the Nekhoroshev theorem sketched in this section is described with all mathematical details in Pöschel (1993). The reader can consult the papers by Nekhoroshev (1977), Benettin *et al.* (1985). and Lochak (1992) for alternative approaches.

6.2.2 Nekhoroshev structure

The construction of the Nekhoroshev theorem can be iterated in order to explore the dynamical structure of the system also for what concerns resonances of order larger than Nekhoroshev's threshold $K \sim 1/\varepsilon^b$ (Morbidelli and Giorgilli, 1995b).

Indeed, in the nonresonant domain defined above (denoted hereafter by $\mathcal{G}_K$), after the elimination of the resonances up to order K, in the new action–

Table 6.1: Scheme of iteration of the Nekhoroshev theorem. The first column gives the number of iterations N; the second the size ϱ of the connected components of the domain of definition of the Hamiltonian; the third the size of the perturbation $\|\mathcal{H}_1\|$; the fourth the cut-off order K of the considered resonances; the fifth the Nekhoroshev stability time T_S; the sixth the size ϱ_K of each connected component of the nonresonant domain; and the last one gives the size of the remainder $\mathcal{R}_K$ of the normal form constructed on the nonresonant domain. The iteration consists in applying the Nekhoroshev theorem subsequently on the nonresonant domain defined at the previous step, i.e. substituting columns 6 and 7 into columns 2 and 3. Powers of ε are not indicated

N	ϱ	$\|H_1\|$	K	T_S	ϱ_K	$\|\mathcal{R}_K\|$
1	1	ε	$1/\varepsilon$	$\exp(1/\varepsilon)$	ε	$\exp^{-1}(1/\varepsilon)$
2	ε	$\exp^{-1}(1/\varepsilon)$	$\exp(1/\varepsilon)$	$\exp^2(1/\varepsilon)$	$\exp^{-1}(1/\varepsilon)$	$\exp^{-2}(1/\varepsilon)$
3	$\exp^{-1}(1/\varepsilon)$	$\exp^{-2}(1/\varepsilon)$	$\exp^2(1/\varepsilon)$	$\exp^3(1/\varepsilon)$	$\exp^{-2}(1/\varepsilon)$	$\exp^{-3}(1/\varepsilon)$
...	...	...	...	...	...	...
n	$\exp^{-n+2}(1/\varepsilon)$	$\exp^{-n+1}(1/\varepsilon)$	$\exp^{n-1}(1/\varepsilon)$	$\exp^n(1/\varepsilon)$	$\exp^{-n+1}(1/\varepsilon)$	$\exp^{-n}(1/\varepsilon)$
...	...	...	...	...	...	...

angle variables $\mathbf{p}', \mathbf{q}'$ the Hamiltonian has the form:

$$H'(\mathbf{p}', \mathbf{q}') = \mathcal{H}_0'(\mathbf{p}') + \varepsilon'\mathcal{H}_1'(\mathbf{p}', \mathbf{q}') \ , \quad \text{with } \varepsilon'\mathbf{H}_1'(\mathbf{p}', \mathbf{q}') \equiv \mathcal{R}_K(\mathbf{p}', \mathbf{q}') \ . \quad (6.4)$$

This is again an autonomous convex system, but the new perturbation is exponentially small with respect to the original one, because $\varepsilon' \sim \exp(-1/\varepsilon^b)$. Then, applying the Nekhoroshev theorem to this Hamiltonian (i.e. considering resonances up to a new cut-off $K' \sim 1/\varepsilon'^b \sim \exp[b/\varepsilon^b]$) one proves the global stability of motion in $\mathcal{G}_K$ for superexponentially long times, namely up to $T \sim \exp[\exp(1/\varepsilon^b)]^b$. Moreover, one finds a new nonresonant domain $\mathcal{G}_{K'}$, characterized by the absence of resonances up to order K'. On $\mathcal{G}_{K'}$ one can introduce new action–angle variables so as to transform the Hamiltonian into the sum of an integrable part and a remainder $\mathcal{R}_{K'}$, the latter superexponentially small. This procedure can be iterated, and it is proved to be convergent to a set of invariant KAM tori of large volume (Giorgilli and Morbidelli, 1997). The fact that, at each step of the iteration, the nonresonant domain is fragmented into smaller and smaller pieces never prevents the application of the Nekhoroshev theorem: indeed the most relevant parameter, i.e. the ratio between the size of the perturbation and the size of each connected component of the domain, decreases at each step. The iteration procedure of

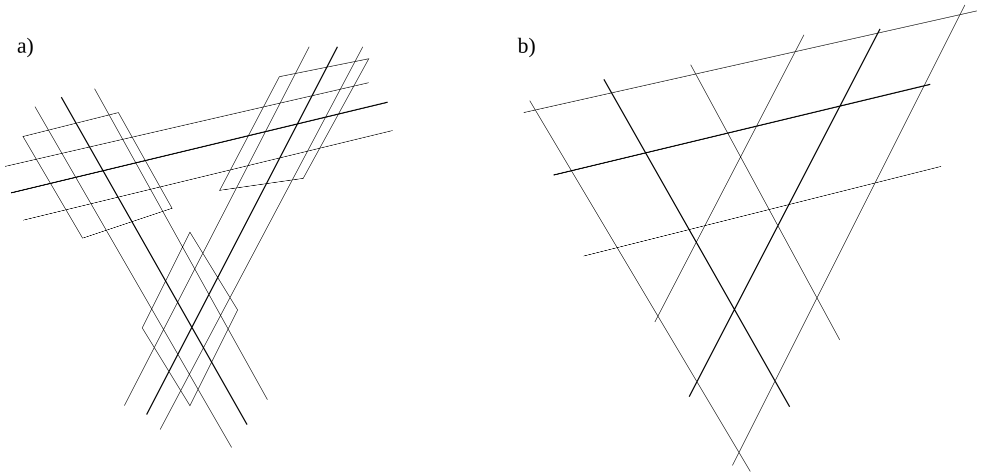


Figure 6.7: Panel a: Sketch of the Nekhoroshev structure. The resonances of a given order cross over at resonant nodes but do not overlap: there is always a nonresonant domain within each mesh of the resonance web. This picture repeats at every scale, i.e. looking subsequently to resonances with increasing order. Panel b: resonances of some order overlap. The nonresonant domain cannot be defined. Invariant tori are destroyed. Motion can pass from one resonance to another. As in Fig. 6.6, the representation of the resonant structure is sketched here on the frequency plane for a three-degree of freedom system. Reprinted from Fig. 3 of Morbidelli and Guzzo (1996), with permission from Kluwer Academic Publishers.

the Nekhoroshev theorem is summarized in Table 6.1.

In conclusion, the global picture that we obtain in the light of the Nekhoroshev theorem and of its successive iterations is the one sketched in Fig. 6.7. When the perturbation parameter ε is *sufficiently small* the Nekhoroshev theorem can be applied. The dynamics is *structured* (Fig. 6.7a). Resonances of a given order cross over at resonant nodes, in correspondence to the resonances of higher multiplicity, in the neighborhood of which one can find well-defined chaos, but resonances do not completely overlap, in the sense that there is always a nonresonant domain in each mesh of the resonance web. Moreover, this scenario is true at every scale, i.e. looking successively at resonances of increasing order. The core of this structure is the set of invariant KAM tori, which in this view is the nonresonant domain in the limit where the number of iterations of the Nekhoroshev theorem is pushed to infinity. We will say hereafter that a dynamical system characterized by such a structure is a

Nekhoroshev system.

The inverse situation is the one sketched in Fig.6.7b. At some order, resonances overlap. There is no place for a nonresonant domain, and invariant tori no longer exist. The overlapping of resonances allows orbits to pass from one resonance to another, in a chaotic "fast" (Chirikov-like) diffusion. The phase space is affected by *large-scale chaos.* Since the width of single resonant domains scales as $\sqrt{\varepsilon}$ this situation happens if ε is not *sufficiently small.*

6.2.3 Superexponential stability of KAM tori

From Table 6.1, comparison between the size ϱ of the nonresonant domain and its stability time T_S, at each order n of the iteration of the Nekhoroshev theorem, suggests that the time required to escape from the vicinity of a KAM torus is superexponentially long with the inverse of the initial distance from the torus.

This can also be proven in a simple direct way (Morbidelli and Giorgilli, 1995a). As shown in Section 3.2, starting from the Kolmogorov normal form, in the neighborhood $\mathcal{U}_\varrho$ of radius ϱ of a KAM torus one can construct an optimal Birkhoff normal form, whose remainder has size $\varepsilon_\varrho \sim \exp[\varrho^{-1/(\tau+1)}]$. The Hamiltonian in $\mathcal{U}_\varrho$ is therefore written as $\mathcal{H}(\mathbf{I},\boldsymbol{\varphi}) = \mathcal{H}_0(\mathbf{I}) + \varepsilon_\varrho \mathcal{R}(\mathbf{I},\boldsymbol{\varphi})$. If $\mathcal{H}_0(\mathbf{I})$ satisfies the hypotheses of the Nekhoroshev theorem, applying the latter one proves that the time required to escape from $\mathcal{U}_\varrho$, starting from $\mathcal{U}_\varrho - \varepsilon_\varrho$, is proportional to $\exp[1/\varepsilon_\varrho^b] = \exp\{\exp[b/\varrho^{1/(\tau+1)}]\}$.

This improves the result on the stability of KAM tori already discussed in Section 3.2, and stresses the importance that KAM tori have in structuring the dynamics of Hamiltonian systems with any finite number of degrees of freedom.

6.3 Exploring the dynamical structure of a given system

At the light of Section 6.2.2, the correct attitude for approaching the investigation of a given dynamical system is to search for an answer to the following questions: has the system the Nekhoroshev structure? In the negative case, at which order do resonances overlap?

Unfortunately, it is not an easy task to answer these questions in an analytic way. This can be understood with the following qualitative argument, inspired by the work of Arnold (1963b). The width of a resonant domain scales as the square root of the coefficient of the corresponding resonant harmonic, i.e. scales with the order K of the resonance as $\sqrt{\varepsilon}\exp(-K\sigma)$; the number

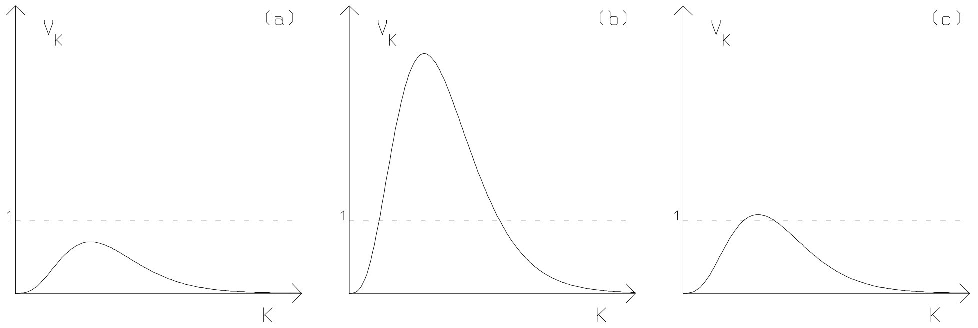


Figure 6.8: The relative volume V_K filled by all resonances of order K, as given by formula (6.5). In the first picture, ε is small, so that for any K the volume V_K is always smaller than the volume of the phase space (here normalized to 1). Therefore resonances cannot globally overlap; invariant tori exist and the system has the Nekhoroshev structure. In the second and third pictures, the volume V_K is larger than 1 for some K. Therefore resonances must overlap, and invariant tori cannot exist. The order at which resonances globally overlap depends on ε. If ε is very large (second picture), the order is small; otherwise (third picture) the order can be very large, close to the one corresponding to the maximum of V_K. Reprinted from Fig. 4 of Morbidelli and Guzzo (1996), with permission from Kluwer Academic Publishers.

of resonances of order K grows as $2^n K^{n-1}$, n being the number of degrees of freedom. Therefore, the relative volume of the phase space covered by all resonances of order K is

$$V_K \sim 2^n \sqrt{\varepsilon} K^{n-1} \exp(-K\sigma) \ . \tag{6.5}$$

The volume V_K has a maximum value for $K = (n-1)/\sigma$, and decays exponentially for larger K. Figure 6.8 plots V_K as a function of K, for different values of ε. If ε is small, the volume covered by resonances of order K is smaller, for every K, than the volume of the full phase space (Fig. 6.8a). This implies that, at all orders, resonances cannot overlap completely, and there is some volume free from resonances where invariant tori can exist; the system has the Nekhoroshev structure.[2] If ε is larger, then for some K the relative volume V_K is equal to one. As a consequence, the resonances of order K must overlap

[2]Note that, from the rigorous mathematical viewpoint, in order to prove the existence of

completely, invariant tori cannot exist and the system loses its Nekhoroshev structure. The minimal order K at which resonances overlap depends on ε. If ε is very large, this can happen at very low order (Fig. 6.8b), but if ε is smaller, this can happen for K closer to $K_{\max} = (n-1)/\sigma$ (Fig. 6.8c). The latter can be very large if the number n of degrees of freedom is large or the parameter of analyticity σ is small. For example, the dynamical system for a realistic description of the motion of an asteroid has at least 10 degrees of freedom, and σ can be estimated to be about 0.1, so that global resonance overlap can be expected at order 100...!

Now, the analytic algorithms, in the limit of implementation imposed by modern technology (algebraic manipulators, seminumerical computations etc.), allow the investigation of the location and the amplitude of resonances only up to a very limited order. Therefore, one can conclude whether a given system has the Nekhoroshev structure only if $K_{\max}$ is small (small number of degrees of freedom, large σ). As a matter of fact, all successful applications of the Chirikov criterion of resonance overlapping (see Contopoulos, 1966; Chirikov, 1979; Wisdom, 1980) have been achieved on two-degree of freedom systems, where good results can be obtained at low to moderate order.

The situation is not so desperate if one turns to the numerical exploration of dynamical systems. The tools illustrated in Chapter 5 allow one to efficiently detect the resonances and the chaotic zones, so that one can efficiently check whether the resonances overlap, or leave enough room for the existence of KAM tori. A nice demonstration of this way to proceed has been given by Laskar (1993), Kaneko and Konishi (1994) and, more recently, Froeschlé *et al.* (2000). Froeschlé *et al.* studied the three-degree of freedom Hamiltonian

$$\mathcal{H} = \frac{p_1^2}{2} + \frac{p_2^2}{2} + p_3 + \varepsilon \left(\frac{1}{\cos q_1 + \cos q_2 + \cos q_3 + 4} \right) . \tag{6.6}$$

The unusual dependence on the angles of (6.6) has been cleverly chosen in order to have, in a compact form, a Hamiltonian whose Fourier expansion has an infinite number of harmonics. To explore the dynamical structure of this system, the authors computed the fast Lyapunov exponent (see Section 5.4.1) for 250,000 orbits, with initial p_1 and p_2 chosen on a regular grid and initial phases of the angles equal to zero. Figure 6.9 shows, with a color code, the value of the FLI as a function of the initial (p_1, p_2). Let us first focus on the top left panel, which concerns the case with $\varepsilon = 0.01$. The discontinuities of the

invariant tori one has to show that the sum of the V_K, for $K \in [1, \infty)$, is smaller than 1; however, since most high-order resonance domains are actually covered by low-order resonance ones, the sum of V_K leads to an overestimate of the global resonance volume, so that it is more intuitive to show in Fig. 6.8 the V_K distribution rather than its integral.

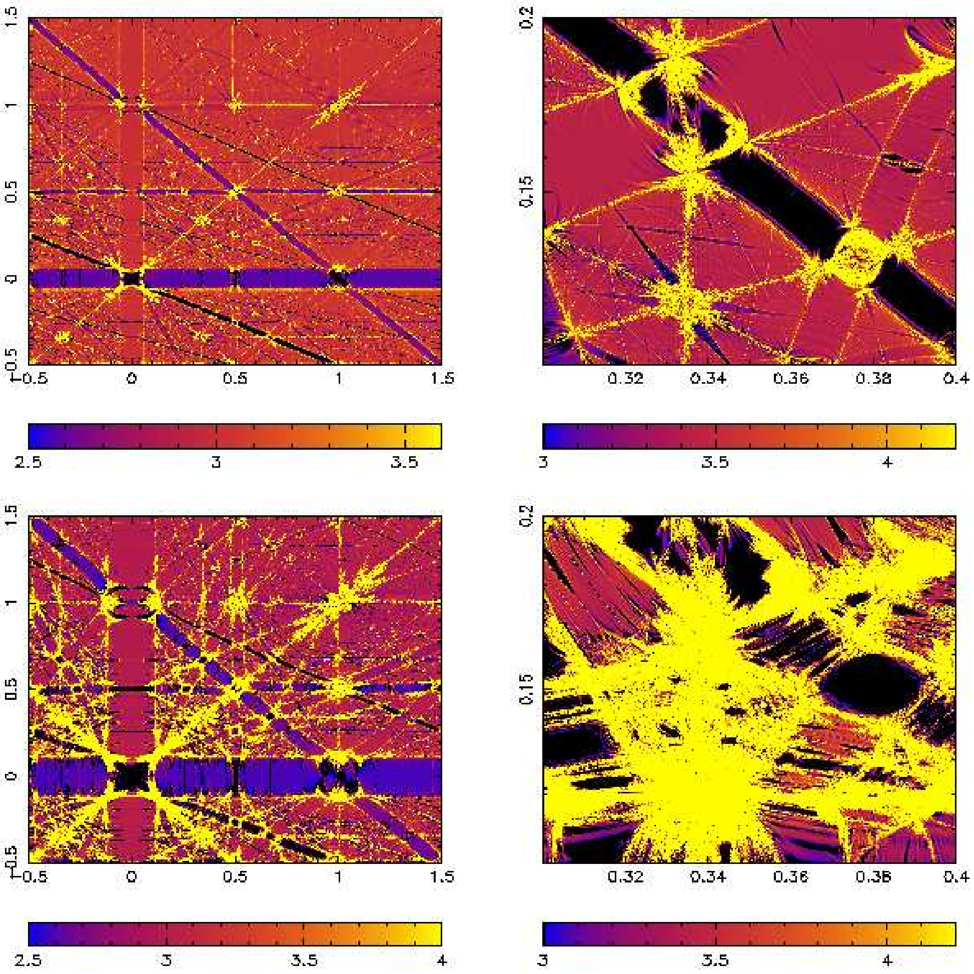


Figure 6.9: The dynamical structure of (6.6) for two values of ε and two different resolutions. The coordinates on each panel are the initial conditions of the actions p_1 and p_2 (the initial values of the angles has been set to zero). The color codes the decimal logarithm of the value of the Fast Lyapunov Indicator, as a function of the initial conditions. 250,000 orbits have been studied in each panel. The top panels refer to the case with $\varepsilon = 0.01$, while the bottom panels are for $\varepsilon = 0.04$. The right panels magnify a region of particular interest of the action plane. The integration time for the computation of the FLI is 2000 on the right panels and 1000 on the left panels. See text for comments. Reprinted from Fig. 2 of Froeschlé *et al.* (2000), with permission from the American Association for the Advancement of Science.

colors relative to the reddish background put in evidence the major resonances of the system. The resonances appear as straight lines or straight bands, because the frequencies of q_1 and q_2 are linear in p_1 and p_2, so that a resonant relationship $k_1\dot{q}_1 + k_2\dot{q}_2 + k_3\dot{q}_3 = 0$ gives a linear equation $k_1p_1 + k_2p_2 + k_3 = 0$ for p_1 and p_2. The resonances which appear as bands are those for which the choice of the initial angles allows the sampling of the regular libration region inside them. In the opposite case, the resonances appear as individual yellow lines. The FLI of the librating orbits is typically smaller than the the FLI of the background nonresonant orbits, so that several bands are colored in blue. The yellow lines (both those bounding the resonant bands and the individual ones) show the separatrices or the chaotic regions associated with the corresponding resonances. In fact, on the separatrices and in the chaotic regions the motion is highly hyperbolic, so that the value of the FLI is large.

The resonant structure shown in the top left panel of Fig. 6.9 recalls the representation of Fig. 6.7a, suggesting that, for this value of ε, the system might have the Nekhoroshev structure. The bottom left panel shows the situation for a value of ε that is four times larger: the resonant bands are wider, and more resonant yellow lines are visible, so that the existence of the Nekhoroshev structure, at least in the region where $|p_1|$ and $|p_2|$ are smaller than 0.5, is more questionable. The real structure of the system appears more clearly if one increases the resolution of the numerical investigation. This requires the construction of a finer grid of initial conditions and to integrate the orbits for a longer timespan, in order to detect weaker chaos and thinner resonances, as done in the right panels of Fig. 6.9. If the system has the Nekhoroshev structure, the situation of nonoverlapping of resonances repeats at every resolution level. Because the relative volume of KAM tori is large, the existence of a nonresonant domain becomes more evident as one increases the resolution of the analysis (top right panel). Conversely, if resonances of some order overlap each other, a sufficiently detailed analysis must reveal the global chaoticity of the system (bottom right panel). Thus, Fig. 6.9 allows the conclusion that, in the considered region of the action space the system (6.6) has the Nekhoroshev structure if $\varepsilon = 0.01$, while it is globally chaotic if $\varepsilon = 0.04$.

This example shows that numerical analysis, coupled with theoretically based interpretations, is a very powerful way to understand the dynamical properties of a given system. In modern Celestial Mechanics this kind of approach is becoming increasingly common (see for instance Chapters 11 and 12).

Chapter 7

SECULAR DYNAMICS OF THE PLANETS

7.1 Lagrange–Laplace solution

With the basic concepts of the theory of Hamiltonian dynamics outlined in the previous chapters, it is now possible to study in detail the dynamics in the Solar System. In this and in the following chapter we study the *secular dynamics*; the latter is the dynamics described by the *secular normal forms* (see Section 2.5.1), once the remainders are neglected. Of course, the secular dynamics is an accurate description of the real dynamics only in those regions where the secular normal form can be constructed, i.e. away from mean motion resonances, and its accuracy is increasingly better with the order of the secular normal form.

For simplicity of notation, in the following we will omit all the superscripts which were used in Section 2.5.1 to stress that the variables appearing in the secular normal form are the *mean* modified Delaunay variables, namely those obtained from the original modified Delaunay variables through the sequence of Lie series. Moreover, we will refer to the mean orbital elements defined from the mean modified Dealunay variables, omitting the adjective *mean*.

The planetary secular normal form does not depend, by definition, on the mean longitudes of the planets $\lambda_1, \ldots, \lambda_N$. As a consequence, the actions $\Lambda_1, \ldots, \Lambda_n$ are constants of motion. The secular system is therefore completely described by the canonical action–angle variables P_j, Q_j, p_j, q_j $(j = 1, \ldots, N)$. Then, the "main" term $\mathcal{H}_0(\Lambda_1, \ldots, \Lambda_n)$ can be dropped from the normal form Hamiltonian, which therefore assumes the structure:

$$\mathcal{H} = \varepsilon \mathcal{H}_1(P_1, \ldots, P_n, Q_1, \ldots, Q_n, p_1, \ldots, p_n, q_1, \ldots, q_n; \Lambda_1, \ldots, \Lambda_n) \ , \qquad (7.1)$$

where for simplicity we have included in $\varepsilon\mathcal{H}_1$ also the terms of the normal form of higher order in ε; the quantities Λ_j play the role of constant parameters.

In (7.1) the small parameter ε (the mass of the largest planet relative to that of the Sun) multiplies the entire Hamiltonian. Therefore, it no longer plays the role of a perturbation parameter separating an integrable part from its perturbation, but simply shows that the motion described by the Hamiltonian (7.1) is *slow*, with a natural timescale that is $1/\varepsilon$ longer than Jupiter's orbital period. Therefore, in order to study the Hamiltonian (7.1) with the tools described in Chapter 2, one first has to find an integrable approximation and a new perturbation parameter – say η – such that, in suitable action–angle variables $\mathbf{p}, \mathbf{q}$, (7.1) can be written as $\mathcal{H}_0(\mathbf{p}) + \mathcal{H}_1(\mathbf{p}, \mathbf{q})$, with $\mathcal{H}_1$ of order η with respect to $\mathcal{H}_0$.

The classical way to proceed is the following. Noting that the eccentricities and the inclinations of the planets are small, and that

$$P_j \sim \sqrt{a_j}\frac{e_j^2}{2} \ , \quad Q_j \sim 2\sqrt{a_j}\sin^2\frac{i_j}{2} \ , \tag{7.2}$$

one expands (7.1) in a Taylor series in $\sqrt{P_j}, \sqrt{Q_j}$, writing

$$\mathcal{H}(P_j, Q_j, p_j, q_j) = \sum_{n\geq 0} \mathcal{H}_{(n)}(P_j, Q_j, p_j, q_j) \tag{7.3}$$

where $\mathcal{H}_{(n)}$ is a polynomial function of $\sqrt{P_j}, \sqrt{Q_j}$ of degree n. The first term $\mathcal{H}_{(0)}$ of the series (7.3) vanishes, because D'Alembert rules (see Section 1.9.3) imply that harmonics of p_j, q_j with coefficients that do not depend on P_j, Q_j cannot exist. Also, all the terms $\mathcal{H}_{(n)}$ with odd n vanish, because the harmonics of the secular problem – whose general form is $\prod_j P_j^{\alpha_j/2} Q_j^{\beta_j/2} \exp[\iota \sum_j (m_j p_j + k_j q_j)]$ – must have $\sum_j (m_j + k_j) = 0$; this implies that $\sum_j (|m_j| + |k_j|)$ is even and therefore that also $\sum_j (\alpha_j + \beta_j)$ is even. As a consequence, the leading term in (7.3) is $\mathcal{H}_{(2)}$. Still because of the D'Alembert rules, the general form of $\mathcal{H}_{(2)}$ must be:

$$\mathcal{H}_{(2)} = \sum_{j=1}^{N}\sum_{k=1}^{N} \left[c_{j,k}\sqrt{2P_j}\sqrt{2P_k}\cos(p_j - p_k) + d_{j,k}\sqrt{2Q_j}\sqrt{2Q_k}\cos(q_j - q_k)\right] \ , \tag{7.4}$$

the coefficients $c_{j,k}, d_{j,k}$ depending only on the constants Λ_j, namely on the semimajor axes of the planets. Note that when $k = j$ the terms in the sum become $2c_{j,j}P_j + 2d_{j,j}Q_j$. Introducing now the canonical polynomial variables (1.78), (7.4) becomes:

$$\mathcal{H}_{(2)} = \sum_{j=1}^{N}\sum_{k=1}^{N} [c_{j,k}(x_j x_k + y_j y_k) + d_{j,k}(v_j v_k + z_j z_k)] \ , \tag{7.5}$$

i.e. a quadratic polynomial form. The equations of motion of (7.5) are therefore linear and – as all linear differential equations – they are integrable. Thus, we have identified in $\mathcal{H}_{(2)}$ an integrable approximation of the secular Hamiltonian (7.1). The terms $\mathcal{H}_{(n)}$ with $n \geq 4$ will play the role of a perturbation, their size relative to $\mathcal{H}_{(2)}$ being $[\max_j(\sqrt{P_j}, \sqrt{Q_j})]^{(n-2)} \sim [\max_j(e_j, i_j)]^{(n-2)}$. As a consequence, while in the original planetary problem (2.39) the natural perturbation parameter ε is the largest planetary mass relative to that of the Sun, in the secular problem the natural perturbation parameter η becomes the square of the largest value assumed by the planetary eccentricities or inclinations during the secular evolution.

We now proceed to compute the solution of the equations of motion given by $\mathcal{H}_{(2)}$ – the so-called *Lagrange–Laplace solution* – and to introduce suitable action–angle variables that allow one to write $\mathcal{H}_{(2)}$ as a function of the sole actions. For this purpose, we first look for new canonical variables x'_j, y'_j, v'_j, z'_j such that (7.5) can be written as:

$$\mathcal{H}_{(2)} = -\sum_j \left[g_j \frac{(x'_j)^2 + (y'_j)^2}{2} + s_j \frac{(v'_j)^2 + (z'_j)^2}{2} \right] , \tag{7.6}$$

where g_j and s_j are coefficients depending only on the mean semimajor axes of the planets. Then it will be easy to introduce new action–angle variables P'_j, p'_j, Q'_j, q'_j with the inverse of transformation (1.78), to write

$$\mathcal{H}_{(2)} = -\sum_j [g_j P'_j + s_j Q'_j] . \tag{7.7}$$

The Hamiltonian (7.5) can be rewritten in matrix form as:

$$\mathcal{H}_{(2)} = \mathbf{x} \cdot A\mathbf{x} + \mathbf{y} \cdot A\mathbf{y} + \mathbf{v} \cdot B\mathbf{v} + \mathbf{z} \cdot B\mathbf{z}, \tag{7.8}$$

where A and B are symmetric $N \times N$ matrices. Thus, the transformation of (7.5) into (7.6) is reduced to the simple diagonalization of the matrices A and B. The latter being symmetric, the diagonalization can be achieved with a rotation of the vectors. The transformation to the new variables is therefore

$$\mathbf{x} = R^I \mathbf{x}' , \quad \mathbf{y} = R^I \mathbf{y}' , \quad \mathbf{v} = R^{II} \mathbf{v}' , \quad \mathbf{z} = R^{II} \mathbf{z}' , \tag{7.9}$$

where R^I and R^{II} are $N \times N$ rotation matrices and $\mathbf{x}, \mathbf{y}, \mathbf{v}, \mathbf{z}$ respectively denote the vectors with components $(x_1, \ldots, x_N)$, $(y_1, \ldots, y_N)$, $(v_1, \ldots, v_N)$, $(z_1, \ldots, z_N)$. We finally prove that (7.9) is a canonical transformation: a linear transformation on the coordinates $\mathbf{x} = C\mathbf{x}'$ is canonical if and only if the transformation on the momenta has the form $\mathbf{y} = (C^T)^{-1}\mathbf{y}'$, where

$(C^T)^{-1}$ denotes the inverse of the transposition of the matrix C (this can be easily checked directly on the Hamilton equations); in (7.9) the coordinates and the conjugate momenta are transformed with the *same* rotation matrix, but rotation matrices R have precisely the general property that $R = (R^T)^{-1}$.

Once the action–angle variables P'_j, p'_j, Q'_j, q'_j are introduced and $\mathcal{H}_{(2)}$ is transformed into (7.7), it becomes evident that, if one neglects the perturbation $\sum_{n\geq 2} \mathcal{H}_{(2n)}$, the actions P'_j and Q'_j are constants of motion and the angles p'_j and q'_j move linearly with time, with fixed frequencies $-g_j$ and $-s_j$ respectively. By the inverse composition of the performed canonical transformations, one easily computes the motion of the original action–angle variables P_j, p_j, Q_j, q_j as a function of time. Then, recalling the relationships (7.2), the time dependence of the eccentricities, inclinations, longitudes of perihelion and node of the planets are:

$$\begin{aligned} e_j \cos \varpi_j &= \textstyle\sum_{k=1}^N M_{j,k} \cos(g_k t + \beta_k) \\ e_j \sin \varpi_j &= \textstyle\sum_{k=1}^N M_{j,k} \sin(g_k t + \beta_k) \\ \sin \tfrac{i_j}{2} \cos \Omega_j &= \textstyle\sum_{k=1}^N N_{j,k} \cos(s_k t + \delta_k) \\ \sin \tfrac{i_j}{2} \sin \Omega_j &= \textstyle\sum_{k=1}^N N_{j,k} \sin(s_k t + \delta_k)\,. \end{aligned} \tag{7.10}$$

This is usually known as the *Lagrange–Laplace solution* for secular planetary motion. The frequencies g_k, s_k, the coefficients $M_{j,k}, N_{j,k}$ and the phases β_k, δ_k, depend at this level – i.e. considering only the $\mathcal{H}_{(2)}$ part of the secular Hamiltonian – only on the semimajor axes of the planets, but will become functions also of P'_j, Q'_j when the perturbation terms $\mathcal{H}_{(2n)}$ with $n \geq 2$ are taken into account (see next section). Accurate values will be listed in Tables 7.1, 7.2 and 7.3 below.

One of the nodal frequencies $s_1, \ldots, s_N$ is necessarily equal to zero, as a consequence of the conservation of total angular momentum. In fact, choosing an arbitrary orthogonal reference frame, the x and y components of the total angular momentum vector are:

$$\begin{aligned} C_x &= \textstyle\sum_{j=1}^N \mu_j \sqrt{\mathcal{G}(m_0 + m_j) a_j (1 - e_j^2)} \sin i_j \cos \Omega_j \ , \\ C_y &= \textstyle\sum_{j=1}^N \mu_j \sqrt{\mathcal{G}(m_0 + m_j) a_j (1 - e_j^2)} \sin i_j \sin \Omega_j \ ; \end{aligned} \tag{7.11}$$

C_x and C_y being constants, one of the longitudes of node can be expressed as a function of the orbital elements of the other planets, namely:

$$\tan \Omega_{\bar{j}} = \frac{C_y - \sum_{j \neq \bar{j}} \mu_j \sqrt{\mathcal{G}(m_0 + m_j) a_j (1 - e_j^2)} \sin i_j \sin \Omega_j}{C_x - \sum_{j \neq \bar{j}} \mu_j \sqrt{\mathcal{G}(m_0 + m_j) a_j (1 - e_j^2)} \cos i_j \sin \Omega_j}\ . \tag{7.12}$$

Therefore, the system (7.7) can have only $N-1$ independent frequencies s_k. By convention, the null frequency is denoted by s_5. Therefore, the action Q'_5 disappears from the expression (7.7) of $\mathcal{H}_{(2)}$. Moreover, every term $\mathcal{H}_{(2n)}$ in the expansion (7.3) must be independent of Q'_5; in fact, if Q'_5 appeared in one of the terms – say $\mathcal{H}_{(2\bar{n})}$ – then a new independent frequency $\dot{q}'_5 = \partial\mathcal{H}_{(2\bar{n})}/\partial Q'_5$ would appear at order $\bar{n}$, thus violating the conservation of angular momentum.

If the reference frame is chosen such that the x, y plane is orthogonal to the total angular momentum vector, the constants $N_{j,5}$ are equal to zero for all j. In this case, the x, y plane is called the *invariant plane*. Otherwise, the constants $N_{j,5}$ are all equal to a value related to the inclination of the x, y plane relative to the invariant plane. All the other constants $M_{j,k}$ (for all k) and $N_{j,k}$ (for $k \neq 5$) do not depend on the choice of the reference frame. Note that from (7.10) if $N_{j,5}$ is zero, all the nodes Ω_j circulate; conversely, if the reference frame is chosen sufficiently inclined with respect to the invariant plane, the nodes Ω_j appear to librate. This is precisely the case of Jupiter's and Saturn's nodes, if one chooses the ecliptic reference frame (the x, y plane coincident with the present Earth's orbital plane).

In the jargon of Celestial Mechanics, the frequencies g_k, s_k are often called the *proper frequencies of the k-th planet.* In reality, the orbital elements of each planet depend on all the frequencies $g_1, \ldots, g_N$, $s_1, \ldots, s_N$, as formula (7.10) shows. The jargon is motivated by the fact that in most of the cases g_k and s_k are the frequencies with the largest amplitudes in the motion of the k-th planet, i.e. $|M_{k,k}| = \max_j |M_{k,j}|$ and $|N_{k,k}| = \max_j |N_{k,j}|$. However this is not true in all cases, as Table 7.2 shows ($M_{7,5}$ is the maximum of all $M_{7,j}$, i.e. the leading frequency of Uranus' perihelion is g_5).

7.2 Higher-order solutions

In the new action–angle variables P'_j, p'_j, Q'_j, q'_j the secular Hamiltonian is ready to be studied with the tools discussed in Chapter 2. For this purpose, the Hamiltonian $\mathcal{H}_{(2)}(P'_j, Q'_j)$ and its perturbation $\sum_{n>1}\mathcal{H}_{(2n)}(P'_j, Q'_j, p'_j, q'_j)$ must be respectively identified with $\mathcal{H}_0(\mathbf{p})$ and $\varepsilon\mathcal{H}_1(\mathbf{p}, \mathbf{q})$ in (2.1). The value of the perturbation parameter ε is here equal to η, namely the square of the largest value assumed by the planetary eccentricities or inclinations during the secular evolution. In this case, it is not necessary to introduce a cut-off K on the order of the harmonics to divide the perturbation as in (2.19), the functions $\mathcal{H}_{(2n)}$ with $n > 1$ already providing a natural partition of $\mathcal{H}_1$ in terms of increasingly higher order with respect to the perturbation parameter. Each function $\mathcal{H}_{(2n)}$ contains only a finite number of harmonics, due to the D'Alembert rules.

To study the secular evolution of the planets it is convenient to construct

the Birkhoff normal form.[1] As shown in Chapter 2, at order $\bar{n}$, the Birkhoff normal form is of type:

$$\mathcal{H}^{\bar{n}} = \sum_{n=1}^{\bar{n}+1} \bar{\mathcal{H}}^{\bar{n}}_{(2n)}(P^{\bar{n}}_j, Q^{\bar{n}}_j) + \mathcal{R}(P^{\bar{n}}_j, Q^{\bar{n}}_j, p^{\bar{n}}_j, q^{\bar{n}}_j) \ , \tag{7.13}$$

where $P^{\bar{n}}_j, Q^{\bar{n}}_j, p^{\bar{n}}_j, q^{\bar{n}}_j$ are the new action–angle variables introduced through a sequence of $\bar{n}$ Lie transforms, and the remainder $\mathcal{R}$ contains terms that are at least of order $\bar{n}+2$ in $P^{\bar{n}}_j, Q^{\bar{n}}_j$.

Neglecting the remainder term, the solution of (7.13) is simply

$$\begin{aligned} P^{\bar{n}}_j(t) = P^{\bar{n}}_j(0) \ , \quad p^{\bar{n}}_j(t) &= \frac{\partial \left(\sum_{n=1}^{\bar{n}+1} \bar{\mathcal{H}}^{\bar{n}}_{(2n)} \right)}{\partial P^{\bar{n}}_j} t + p^{\bar{n}}_j(0) \\ Q^{\bar{n}}_j(t) = Q^{\bar{n}}_j(0) \ , \quad q^{\bar{n}}_j(t) &= \frac{\partial \left(\sum_{n=1}^{\bar{n}+1} \bar{\mathcal{H}}^{\bar{n}}_{(2n)} \right)}{\partial Q^{\bar{n}}_j} t + q^{\bar{n}}_j(0) \ . \end{aligned} \tag{7.14}$$

The motion depends on the initial values $P^{\bar{n}}_j(0), Q^{\bar{n}}_j(0), p^{\bar{n}}_j(0), q^{\bar{n}}_j(0)$; the latter are computed, by inversion of the Lie series, from the initial values of P'_j, Q'_j, p'_j, q'_j, which in turn depend on the initial P_j, Q_j, p_j, q_j through the transformations described in the previous section. Once their initial values are computed, the motion of $P^{\bar{n}}_j, Q^{\bar{n}}_j, p^{\bar{n}}_j, q^{\bar{n}}_j$ is completely determined, and its image in the original variables can be simply computed by applying again (in reverse order) all the canonical transformations. At the end of this chain of transformations, the time dependence of $e_j \cos \varpi_j$, $e_j \sin \varpi_j$, $\sin i_j/2 \cos \Omega_j$, $\sin i_j/2 \sin \Omega_j$ turns out to be the sum of two parts. The first part has the same functional form as the right hand side of (7.10), but with modified values of the coefficients g_k, s_k, β_k, δ_k, $M_{j,k}$, $N_{j,k}$. For instance, the frequencies g_k and s_k are modified by respectively $-\partial\left(\sum_{n=2}^{\bar{n}} \bar{\mathcal{H}}^{\bar{n}+1}_{(2n)}\right)/\partial P^{\bar{n}}_k(P^{\bar{n}}_j(0), Q^{\bar{n}}(0))$ and $-\partial\left(\sum_{n=2}^{\bar{n}+1} \bar{\mathcal{H}}^{\bar{n}}_{(2n)}\right)/\partial Q^{\bar{n}}_k(P^{\bar{n}}_j(0), Q^{\bar{n}}(0))$; both these terms depend on $P^{\bar{n}}_j(0), Q^{\bar{n}}_j(0)$. Note however that, because none of the terms $\mathcal{H}_{(2n)}$ depends on Q'_5, q'_5, the values of s_5 and $N_{j,5}$ are unchanged. The second part contains new harmonics that involve integer linear combinations of the frequencies g_k, s_k. More specifically, the Birkhoff normal form at order $\bar{n}$ introduces harmonics involving combinations of $2\bar{n}+1$ frequencies.

[1] With respect to what was explained in Chapter 2, in this case there is the technical complication that the functions $\mathcal{H}_{(2n)}$ are not analytic when one of the actions is equal to zero, because of the square root terms. However, the D'Alembert rules ensure that the results of the Poisson brackets of the Lie series can always be expanded in positive power series of the square roots of the action variables.

Table 7.1: Planetary frequencies and angular phases. The values of $g_1, \ldots, g_4$ and $s_1, \ldots, s_4$ are taken from Laskar (1990); those of $g_5, \ldots, g_8$ and $s_6, \ldots, s_8$ are taken from Nobili *et al.* (1989); frequencies are measured in arcsec/year. The phases $\beta_1, \ldots, \beta_8$ and $\delta_1, \ldots, \delta_8$ are taken from Laskar (1990) and are computed with respect to the ecliptic reference frame of the year 2000 at the epoch JD 2451545.0.

k	g_k	$\beta_k(°)$	s_k	$\delta_k(°)$
1	5.5964	112.08	−5.6174	348.60
2	7.4559	200.51	−7.0795	273.25
3	17.3646	305.12	−18.8512	240.20
4	17.9156	335.38	−17.7482	303.75
5	4.2575	30.65	0.0000	107.58
6	28.2455	128.09	−26.3450	307.29
7	3.0868	121.36	−2.9927	320.62
8	0.6726	74.06	−0.6925	203.90

The most accurate analytic theory of the secular motion of the eight planets of the Solar System is that of Bretagnon (1974). Actually, Bretagnon did not develop his theory using the Hamilton equations, but his work is equivalent to computing the secular normal form up to order 2 (i.e. ε^2) in the planetary masses, and then computing the Birkhoff normal form (7.13) of order $\bar{n} = 1$, i.e. taking into account in the equations of motion terms of order two ($\mathcal{H}_{(2)}$) and four ($\mathcal{H}_{(4)}$) in the planetary eccentricities and inclinations. Bretagnon later (1984) developed a more accurate theory starting from a model that included also relativistic effects and the perturbations of the Moon on the inner planets.

Duriez (1979) developed an analytic theory of the secular motion of the outer planets (Jupiter to Neptune); his approach is equivalent to computing the Birkhoff normal form up to order $\bar{n} = 2$.

Laskar (1985) computed the secular normal form up to order 2 in the planetary masses and degree 6 in eccentricities and inclinations for the full planetary system (Mercury to Neptune), but then, instead of studying it analytically,

Table 7.2: Coefficients $M_{j,k}$ of the Lagrange–Laplace type solution (7.10). The values with $i \geq 5$ and $k \geq 5$ are taken from Nobili *et al.* (1989); the others are taken from Laskar (1990), when reported, or Bretagnon (1974) otherwise. All values have been multiplied by 10^6.

$j \backslash k$	1	2	3	4	5	6	7	8
1	185444	−27700	1458	−1428	36353	113	623	7
2	6668	20733	−11671	13464	19636	−551	614	11
3	4248	16047	9406	−13159	18913	1506	650	12
4	650	2917	40133	49032	20300	7030	862	20
5	−7	−12	−1	0	44187	−15700	1814	58
6	−6	−12	−7	−7	32958	48209	1511	57
7	2	3	0	0	−37587	−1547	29033	1666
8	0	0	0	0	1881	−103	−3697	9118

resorted to numerical integrations to calculate the frequencies of the motion (Laskar, 1988), on a time interval of 10 My. In principle, the frequencies computed by Laskar are more accurate than those computed by Bretagnon, because he started from a secular normal form containing more perturbation terms and of which numerical integration gave the 'exact' solution.

The most accurate computation of the secular frequencies of the outer planets have been achieved by numerically integrating the full equations of motion (i.e. avoiding the computation of the secular normal form) by Applegate *et al.* (1986) and Nobili *et al.* (1989). With respect to Laskar's computation, the largest discrepancy occurs for the g_6 frequency; this discrepancy is believed to arise from terms of order 3 in the planetary masses, which are missing in Laskar's secular normal form.

In Tables 7.1, 7.2 and 7.3 we list what should be the most accurate determinations of the constants entering Lagrange–Laplace formulæ (7.10). The complete time dependence of the planetary orbital elements, including also the coefficients of the harmonics involving linear combinations of the frequencies, can be found in Bretagnon (1974), Laskar (1988, 1990), Nobili *et al.* (1989).

Table 7.3: The same as Table 7.2, but for the coefficients $N_{j,k}$. The ecliptic reference frame for the year 2000 has been used for the computation of $N_{j,5}$.

$j \backslash k$	1	2	3	4	5	6	7	8
1	39957	30169	1678	72261	13772	−139	−1665	−724
2	6716	−4045	−9544	−5759	13772	−60	−959	−663
3	4960	−3431	8760	4024	13772	−1404	−866	−650
4	860	−566	−15421	34689	13772	−4579	−628	−615
5	−11	4	0	−1	13772	3153	−485	−584
6	−14	6	−2	−13	13772	−7858	−394	−564
7	11	−3	0	1	13772	353	8887	543
8	0	0	0	0	13772	38	−1062	5790

7.3 Chaotic secular motion of the planets

To study the long-term behavior of the secular evolution of the planets, Laskar (1989) numerically integrated for 200 My the equations of motion of his secular normal form. In order to have an accurate representation of the "real" motion, he first made empirical corrections to the frequencies g_6, g_7 and s_7 in $\mathcal{H}_{(2)}$, matching the values measured in the numerical integrations of the full equations of motion of the outer planets. Laskar's numerical solution of the secular equations has been later tested against a 3 My integration of the full equations of motion of the entire Solar System done by Quinn *et al.* (1991), which confirmed its excellent accuracy (Laskar *et al.*, 1992b).

Together with the orbital evolution, Laskar (1989) also computed the largest Lyapunov exponent, which turned out to be $\sim 1/5$ My^{-1}, indicating, as discussed in Chapter 5, that the orbital evolution is chaotic. To confirm the chaotic nature of the secular dynamics of the Solar System, Laskar (1990) looked for the resonances responsible for the origin of chaos. For this purpose, he monitored different combinations, compatible with the D'Alembert rules, of the angles p'_j, q'_j defined in Section 7.1, looking for those in libration, or alternating between libration and circulation. Referring to Chapter 4, this corresponds to looking for the critical angles φ_1 (see formula 4.2) of the exist-

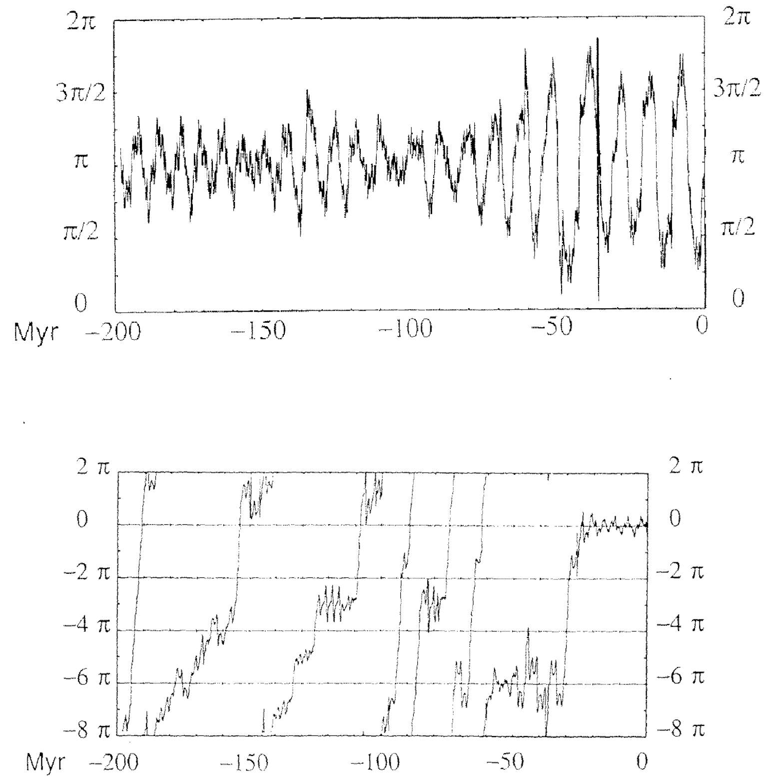


Figure 7.1: The evolution of the angles $p_1'-p_5'-q_1'+q_2'$ (top) and $2(p_4'-p_3')-q_4'+q_3'$ (bottom) according to Laskar's integration of the secular planetary system. The behavior of these angles reveals that the Solar System is in a secular resonance of multiplicity 2, thus explaining the chaotic nature of its secular dynamics. Reprinted from Figs. 2 and 4 of Laskar (1990), with permission from Academic Press.

ing resonances. Laskar found two independent resonant combinations of the secular angles, showing that the Solar System is at least in a resonance of multiplicity 2 (see Chapter 1, Chapter 6). The resonant angles are $p_1'-p_5'-q_1'+q_2'$, whose corresponding harmonic appears in $\mathcal{H}_{(4)}$, and $2(p_4'-p_3')-q_4'+q_3'$, whose corresponding harmonic appears in $\mathcal{H}_{(6)}$, namely respectively at order 4 and 6 in the planetary eccentricities and inclinations. Figure 7.1 shows the time evolution of these two resonant angles. The first one librates during the entire

integration time;[2] note however that the libration amplitude changes significantly and in an irregular manner, showing that the motion is not on an invariant curve of libration, but rather in a chaotic zone, so that transitions to the circulation regime could be expected. In fact these transitions occur in a longer timespan integration in Laskar (1992). The second resonant angle alternates between libration and circulation, showing the genuine chaotic nature of its motion. Laskar also showed that the linear combinations of the actions P'_j, Q'_j, canonically conjugate to the two resonant angles, show oscillations in phases with the librations/circulations of the latter, revealing a pendulum-like dynamics (actually a very confused one, because of the presence of a double resonance). The chaos of the Solar System and the behaviors of the resonant angles has been confirmed through direct numerical integration by Sussman and Wisdom (1992).

As a further confirmation of the existence of chaos, Laskar applied his frequency analysis (see Section 5.3) to monitor the time evolution of the frequencies g_k, s_k of the Solar System. While the numerically measured values of the frequencies $g_5, \ldots, g_8$ and $s_6, \ldots, s_8$ do not have relative changes larger than 10^{-4} over 200 My, those of the frequencies $g_1, \ldots, g_4$ and $s_1, \ldots, s_4$ typically have a relative change of the order of several 10^{-2} (Laskar, 1990). This shows that chaos acts basically only on the eight-degree of freedom subsystem characterizing the secular motion of the inner planets (Mercury to Mars), while the subsystem related to the evolution of the outer planets, being weakly coupled to the first one (the perturbation of the inner planets on the motion of the outer ones is of order of the mass of the Earth relative to the Sun), shows only very limited chaotic activity. This led Laskar to conjecture that the outer planets would evolve on KAM tori if the terrestrial planets did not exist. If this is very likely to be true for the secular system, it is not the case when the nonaveraged system is considered, because of the presence of high-order three-body mean motion resonances (see Chapter 10).

The work of Laskar (1990), however, left open the question about the dynamical structure of the secular planetary problem. In the light of Chapter 6, in fact, it is natural to ask whether the secular system has the Nekhoroshev structure, the chaotic motion being bounded for times exceeding the age of the Solar System. To give a partial answer to this question, Laskar (1994) integrated the equations of motion of his secular normal form for 25 Gy (backwards from 0 to -10 Gy and forward from 0 to 15 Gy). Because of the existence of chaos, which makes the motion unpredictable on timescales of several times

[2] Laskar's integration was a backwards integration, from the present epoch to -200 My; Hamiltonian dynamics being reversible, backwards integrations and forward integrations are perfectly equivalent.

the inverse of the Lyapunov exponent (5 My), the results of these integrations should be regarded only as representative of *possible* evolutions; however, as seen in the simple numerical examples of Chapter 4, numerical integration may give an indication of the extent of the chaotic region that is reachable over the integration time. Note moreover that the integrated timespan exceeds the physical lifetime of the Solar System, which presumably formed ~ 4.5 Gy ago and which should last only 5 billion years more, before the Sun evolves to a red giant star state, and extends its radius up to the Earth's orbit. The long timespan was intended to better reveal the dynamical character of the Hamiltonian secular system. For instance, if a sudden increase of the eccentricity of a planet occurs after 10 Gy, it is plausible that such an event might happen also on a somewhat shorter timespan, for instance within 5 Gy from now. Similarly, what the integration shows to happen at negative times could happen as well in the future.

In order to clearly show the chaotic evolution of each planet and its "diffusion" in the chaotic zone, Laskar adopted the "maximal action method" (see Section 5.4.4), computing the maximal eccentricity and inclination attained by each planet during intervals of 10 My. Figure 7.2 shows that the maximal eccentricity and inclination of the outer planets is constant with time. As discussed in Section 5.4.4, this reveals that the secular dynamics of the outer planets is basically regular. The Earth and Venus show coupled irregular variations of their maximal eccentricities and inclinations, but their values are confined within a narrow band, showing that, despite their chaoticity, their orbits should not change much over the age of the Solar System. The same is not true for Mars and especially for Mercury. Performing a series of numerical integrations with slightly different initial conditions, Laskar showed that Mercury could intersect the orbit of Venus, potentially colliding with the latter or being ejected from the Solar System by a series of close encounters. This result shows that the Solar System is only *marginally stable*, namely that there is a small, but nonnull, probability that a dramatic event, such as the collision between two planets, occurs during its physical lifetime. However, it should be noted that Laskar's equations of secular motion, being obtained through a truncated power expansion in eccentricities and inclinations, could in principle lose accuracy when large values of the latter are achieved.

The marginal stability of the Solar System fits with our present understanding of how the Solar System formed (see Taylor, 1999 for a review). The aggregation of solid material should have led not only to the formation of the present planets, but also of several other planetary embryos with masses that could have reached (especially in the outer Solar System) the Earth's mass. This view is supported by modern numerical simulations of planetary

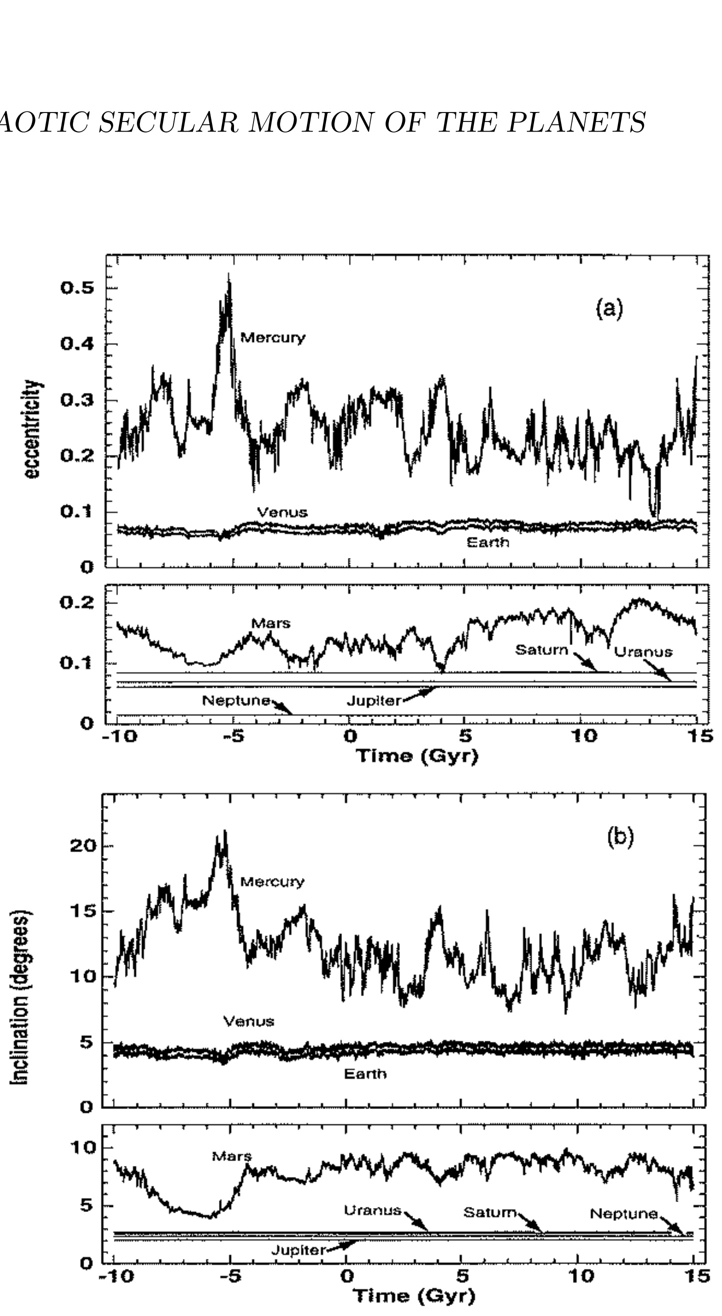


Figure 7.2: The long-term secular evolution of the solar system. The maximal eccentricity and inclination attained by each planet during intervals of 10 My are plotted over a 25 Gy timespan. Preservation of these quantities (as is the case for the giant planets) reveals regular motion, while temporal variations show evidence of chaotic evolution. Reprinted from Fig. 1 of Laskar (1994), with permission from Astronomy and Astrophysics.

formation (Chambers and Wetherill, 1998; Agnor *et al.*, 1999) as well as by "evidence" such as the existence of the Moon (presumably issued from the collision of a martian-sized embryo with the Earth; see Hartman and Davis, 1975; Cameron, 1997) and the obliquity of Uranus (which requires the impact of an Earth-sized body; Safronov, 1966; Parisi and Brunini, 1996). A solar system crowded with massive embryos is violently unstable (Chambers *et al.*, 1996) and evolves, through mutual collisions of embryos or ejections of bodies on hyperbolic orbits, towards a system characterized by a smaller number of embryos and a consequent reduced instability. This kind of evolution effectively "stops" when a marginally stable system is achieved. In this scenario, Mercury would be the last embryo in the Solar System, which hasn't yet had its chance to escape from the system or to collide with another planet.

7.4 Spin axes dynamics

An exhaustive description of the dynamics of the spin axes of the planets goes beyond the purposes of this book. The reader can consult Kinoshita and Souchay (1990) and Souchay *et al.* (1999) for detailed and accurate studies. We discuss here a simplified, but nevertheless highly interesting, representation of the dynamics of the spin axis of a planet that is achieved when a number of approximations are introduced, namely:

1) the planet is assumed to be a rigid homogeneous body with principal momenta of inertia I_1, I_2, I_3 satisfying the relationship $I_1 = I_2 < I_3$; the spin axis is identified with direction of I_3;

2) all direct perturbations induced by the other planets on the rotational motion are neglected; only indirect perturbations are considered, i.e. the changes of the planet's orbit due to the influence of the other planets;

3) the resulting equations are averaged over the orbital periods, which is equivalent to constructing a first-order secular normal form.

In a noninertial orthogonal reference frame x, y, z, with x, y plane coincident with the instantaneous orbital plane of the planet and x axis directed towards the ascending node Ω of the planet's orbit, the orientation of the spin axis can be determined through the angles ε and Λ defined by the relationships

$$s_x = \sin\varepsilon \sin\Lambda \ , \quad s_y = \sin\varepsilon \cos\Lambda \ , \quad s_z = \cos\varepsilon \tag{7.15}$$

where s_x, s_y and s_z are the components in the x, y, z reference frame of the unitary vector directed as the spin axis. The angle ε is usually called the

obliquity of the planet, while the angle $\psi = \Lambda - \Omega$ is called the *longitude* of its spin axis. With the simplifications enumerated above, the quantities $X = \cos\varepsilon$ and ψ turn out to be canonical action–angle variables, whose evolution is described by the one-degree of freedom time-dependent Hamiltonian

$$\mathcal{H}(X, \psi, t) = \frac{\alpha}{2}(1 - e(t))^{-3/2} X^2 + \sqrt{1 - X^2}\,(A(t) \sin\psi + B(t) \cos\psi) \quad (7.16)$$

(Laskar and Robutel, 1993; Laskar *et al.*, 1993). In (7.16) $e(t)$ is the eccentricity of the orbit of the planet, and $A(t), B(t)$ are time-dependent quantities related to the inclination and the node of the planet's orbit relative to its initial orbital plane, namely

$$A(t) = 2(\dot{q}+p(q\dot{p}-p\dot{q}))/\sqrt{1 - p^2 - q^2}\,,\ B(t) = 2(\dot{p}-q(q\dot{p}-p\dot{q}))/\sqrt{1 - p^2 - q^2}\,, \quad (7.17)$$

with $p(t) = \sin[i(t)/2] \sin[\Omega(t)]$, $q(t) = \sin[i(t)/2] \cos[\Omega(t)]$ and $i(0) = 0$. Moreover

$$\alpha = \frac{3\mathcal{G}M_\odot}{2a^3\nu} \frac{I_3 - I_1}{I_1} \quad (7.18)$$

where $M_\odot$ is the mass of the Sun, $\mathcal{G}$ is the gravitational constant, a is the semimajor axis of the planet's orbit (assumed to be fixed) and ν is the spin rate of the planet (the inverse of the planet's rotation period). In the case of the Earth, the influence of the Moon cannot be neglected and, assuming the latter on a circular orbit on the Earth's equatorial plane, a term $(\alpha_M/2)X^2$ needs to be added to the Hamiltonian (7.16), where α_M is given by expression (7.18) with the mass of the Moon replacing that of the Sun and the radius of the Moon's orbit replacing a.

The Hamiltonian (7.16) is a quasi-integrable one. It is the sum of a main integrable part $(\alpha/2)X^2$ – or $(\alpha + \alpha_M)/2X^2$ in the case of the Earth – with two perturbation terms: $(\alpha/2)[(1 - e^2)^{-3/2} - 1]X^2$ and $\sqrt{1 - X^2}\,(A(t) \sin\psi + B(t) \cos\psi)$. The latter are respectively proportional to the square of the eccentricity and to the variation of the inclination relative to its initial value, of the planet's orbit. The integrable part tells us that to a first approximation the obliquity of a planet is constant. The precession rate $\dot{\psi}$ of its spin axis (the so-called *equinox precession*) is proportional to X, so that it is maximal when the obliquity is zero ($X = 1$), it is zero when the obliquity is 90 degrees ($X = 0$, as is approximately the case of Uranus), and it changes sign when the obliquity exceeds 90 degrees (the case of Venus). If the orbit of the planet (and therefore the perturbation terms in (7.16)) are quasi-periodic functions of time with Diophantine frequencies, the KAM and Nekhoroshev theorems can both be applied to the Hamiltonian (7.16), provided the perturbations are small enough (see Section 3.1.1 for the KAM

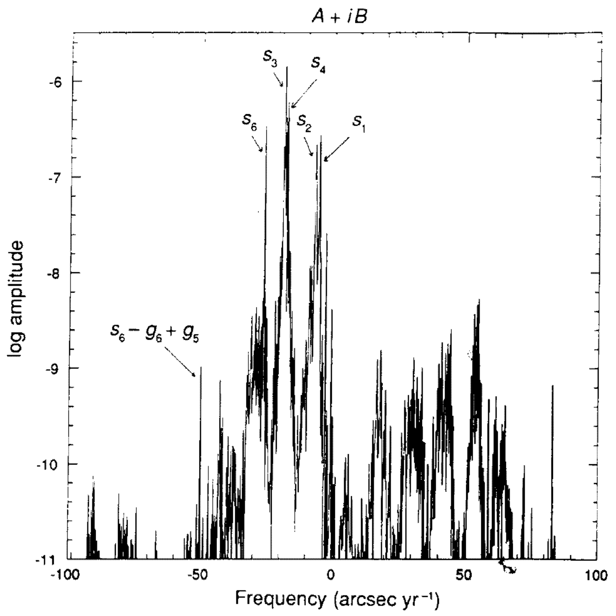


Figure 7.3: The Fourier spectrum of the term $A(t) + \iota B(t)$ entering in (7.16) in the case of the Earth. Reprinted from Fig. 1 of Laskar *et al.* (1993), with permission from Nature, Macmillan Magazines Limited.

theorem and Guzzo and Morbidelli, 1997, for the Nekhoroshev theorem). In particular, if the eccentricity is constant and $A(t) + \iota B(t) = \beta \exp[\iota\omega t]$, (7.16) reduces to the integrable Hamiltonian

$$\mathcal{H} = \frac{\alpha}{2}(1 - e^2)^{-3/2} X^2 + \beta\sqrt{1 - X^2}\sin(\omega t + \psi) \qquad (7.19)$$

which is often referred to as *Colombo's top* from the work of Colombo (1966). The equilibrium points of (7.19) are called the *Cassini states*, and are particularly important for tidally evolving bodies, such as the satellites of Jupiter (see for instance Gladman *et al.*, 1996).

As discussed in the previous section, for the Solar System the quasi-periodic time dependence of the orbital elements is a good approximation only for the outer planets. For instance, Fig. 7.3 shows the Fourier spectrum of $A(t)+\iota B(t)$ for the Earth, according to Laskar's (1990) solution for the secular orbital motion. As one sees, the spectrum is continuous, although strongly peaked around the frequencies of the Lagrange–Laplace linear solution s_1, s_2, s_3, s_4 and s_6 (see Section 7.1). The spectra of $A(t) + \iota B(t)$ for all the other planets are shown in

Laskar and Robutel (1993). Even restricting to the perturbation terms with coefficients exceeding a threshold value, the large number of frequencies makes the Hamiltonian (7.16) very complicated. For this reason, to study the dynamical behavior of the spin axis of the planets of the Solar System, Laskar *et al.* (1993) and Laskar and Robutel (1993) resorted to numerical integrations. They considered a large set of initial conditions $X(0), \psi(0)$, with $\psi(0) = 0$ and initial obliquity $\varepsilon_0 = \cos^{-1} X(0)$ ranging from 0 to 170° in steps of 0.1°; for each initial condition they integrated the evolution for 18 My, following the equations given by (7.16) with $e(t), A(t), B(t)$ from Laskar (1990), and for each evolution they computed the mean frequency of ψ as well as the maximum, minimum and mean value of the obliquity. As discussed in Chapter 5, the dependence of all these quantities on the initial condition reveals the regular or chaotic character of the dynamics.

The left panel of Fig. 7.4 shows the result for the Earth, considering both the torques α and α_M exerted by the Sun and the Moon, and the Earth's present spin rate $\nu = 1/24$ h^{-1}. For initial values of the obliquity smaller than 50° (the Earth's obliquity is 23.3°) the evolution of the spin axis appears very regular, the frequency and the obliquity being smooth functions of the initial ε_0. The difference between the minimal and the maximal value of the obliquity is always very small, showing that the latter has only small-amplitude oscillations. Note that, strictly speaking, the integrated system does not have any regular orbit, because the perturbation is not quasi-periodic and therefore KAM tori cannot exist. Nevertheless, no trace of chaos is detectable, for the following reasons: for the considered values of ε_0, the frequency $\dot{\psi}$ ranges approximately from 35 to 55″/y; because the terms in the perturbation (7.19) are of type $\sin(\psi + \omega t)$, the only possible resonances are those with ω in the range $[-55, -35]''$/y; but the terms in the spectrum of $A + \imath B$ with frequency ω falling in this range have coefficients smaller than 10^{-9}, as can be seen in Fig. 7.3. As a consequence, the amplitude of the chaotic component of the motion of $X(t), \psi(t)$ is so small on the considered timescale to be undetectable with the resolution of Fig. 7.4.

If the initial obliquity is in the interval 55-90°, the evolution of the spin axis is violently chaotic. This is because, on the considered range, the unperturbed frequency $\dot{\psi}$ ranges from 30″/y to 0″/y, and therefore may be resonant with the main terms in the spectrum of $A(t) + \imath B(t)$. The fact that, whatever the initial condition, the obliquity assumes all the values between approximately 55 and 90 degrees over the considered 18 My timespan, suggests that the main resonances, with critical angles respectively equal to $s_6 t + \psi$, $s_3 t + \psi$, $s_4 t + \psi$, $s_2 t + \psi$ and $s_1 t + \psi$, completely overlap each other, giving rise to a uniform chaotic zone and a rapid Chirikov-like diffusion of the action X over all the

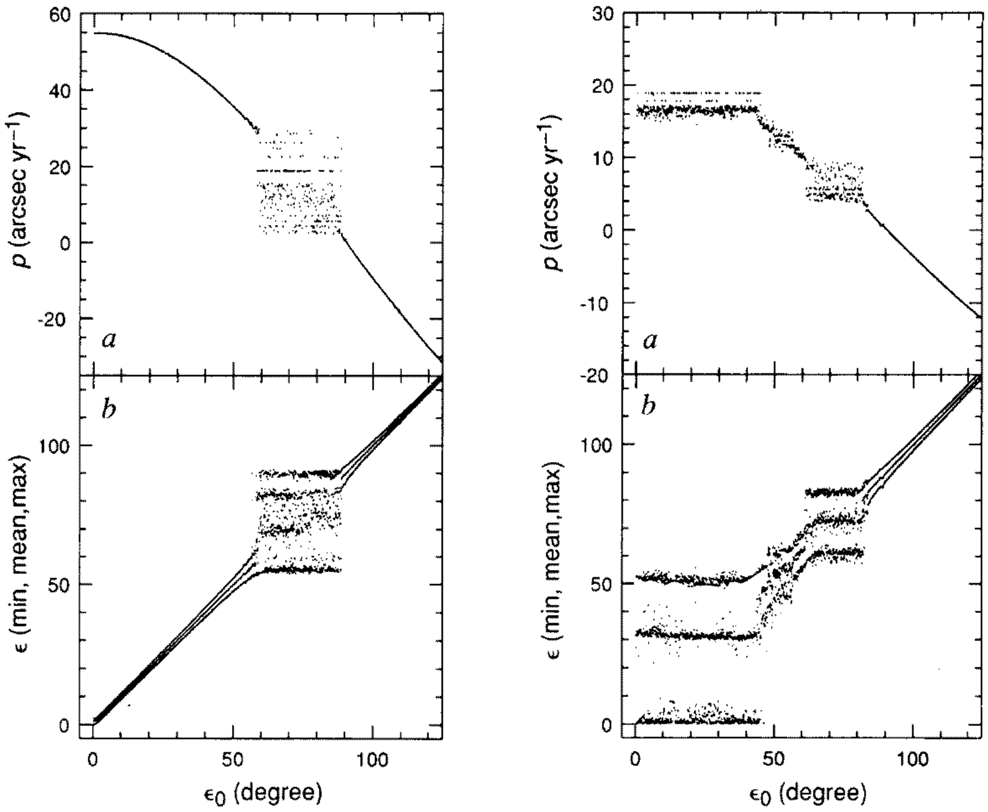


Figure 7.4: The precession frequency (top) and the maximal, mean and minimal obliquity of the Earth's spin axis (bottom) as a function of the initial obliquity ε_0 of the latter. The left panels refer to the model where the torque exerted by the Moon is taken into account, while the right panels concern the case of the Earth without the Moon. See text for discussion. Reprinted from Figs. 2 and 3 of Laskar *et al.* (1993), with permission from Nature, Macmillan Magazines Limited.

resonance overlapping region (see Chapter 6). For obliquities ε_0 exceeding 90°, the motion of the spin axis becomes again basically regular, because none of the terms in the spectrum of $A(t) + \iota B(t)$ with positive frequency has a sufficiently large amplitude.

Laskar *et al.* (1993) also made similar computations to study the dynamics that the Earth's spin axis would have if the Moon did not exist. For this purpose, they suppressed the $\alpha_M X^2$ term in the Hamiltonian; moreover, taking into account that the period of the Earth's rotation has been increased with respect to its primordial value by tidal friction raised by the Moon, they computed α in (7.18) assuming a spin rate ν equal to 1.22 times the present spin rate, in agreement with what is indicated by the geological records for the Earth's rotation frequency 2.5 billion years ago. The right panel of Fig. 7.4 shows the result obtained from this model. For all values of the initial obliquity ε_0 smaller than 80°, the chaotic character of the dynamics is evident. Only obliquities larger than 90° lead to apparent regular motion. Therefore, chaos appears over a larger range of initial obliquities than in the case where the Moon is taken into account. This is simply due to the fact that, for obliquities in the 0-90° interval, the precession frequency ranges only from 0 to 20″/y if the Moon's torque term $\alpha_M X^2$ is not taken into account in the Hamiltonian, while otherwise it ranges from 0 to 55″/y (compare the top plots in Fig. 7.4); on the other hand, quasi-regular motion is possible only for precession frequencies exceeding 30″/y, in order to avoid resonances with the largest amplitude terms in the spectrum of $A(t) + \iota B(t)$.

Notice that, in the case where the Moon is not taken into account, the chaotic region is not uniform. In fact, three regions can be distinguished for initial obliquities respectively in the ranges 0-45°, 45-60° and 60-80°. For ε_0 in the 0-45° region, the range of variation of the obliquity over the integration timespan covers the entire 0-45° interval (notice that the minimal and maximal obliquity do not depend on ε_0). The same situation occurs in the 60-80° region. Conversely, in the 45-60° region the range of variation of the obliquity covers only a small interval around the initial value (so that the minimal and the maximal obliquity are quasi-monotonic functions of ε_0). This implies that the diffusion of the action X is fast in the 0-45° and 60-80° regions while it is slow in the middle region. By looking at the values of the frequency of ψ in the three regions (top plot of the right panel) one realizes that the 0-45° region corresponds to $\dot\psi$ in the range 16-20″/y, and it is probably due to the strong overlapping between the two main resonances with critical angles $s_3 t + \psi$ and $s_4 t + \psi$; similarly, the 60-80° region corresponds to $\dot\psi$ in the range 5-10″/y, and it is probably due to the strong overlapping between the resonances with critical angles $s_1 t+\psi$ and $s_2 t+\psi$; conversely, the 45-60° region corresponds to a

range of $\dot{\psi}$ such that all the resonant terms in the spectrum of $A(t)+\iota B(t)$ have small amplitudes. This region is therefore a "bridge" between the two main chaotic regions; on this "bridge" the diffusion is driven by the overlapping of resonances with smaller widths and it is therefore slower. Crossing the bridge is possible, but only on timescales longer than the one covered by the integrations (18 My).

An interesting question is why the same "bridge" does not happen in the case where the Moon is considered, for which the chaotic region is uniform all over the 0-30$''$/y frequency range. The most plausible answer is the following. We know from Section 4.1 that the width of a resonance in the space of unperturbed frequencies is proportional to $\sqrt{\beta c}$, where c is the coefficient of the resonant harmonic and β is the torque. If the Moon is taken into account, the torque is $\alpha + \alpha_M$, which is greater than the torque α that characterizes the system when the Moon is not considered; moreover in the first case the resonances are located at larger values of the obliquity than in the second case, so that the coefficient of the resonant harmonics c is larger, because it is proportional to $\sqrt{1-X^2}$. Therefore, in the case with the Moon the resonances with critical angles $s_3 t + \psi$ and $s_4 t + \psi$ can overlap with those with critical angles $s_1 t + \psi$ and $s_2 t + \psi$, while in the case without the Moon the resonance widths are not large enough.

Ward (1982) already concluded – with an analytic study of a system equivalent to (7.16) but which included only a few terms in the Fourier expansion of $A(t) + \iota B(t)$ – that without the Moon the obliquity of the Earth would suffer oscillations of $\pm 10°$; the results by Laskar *et al.* (1993) show that the variations of the Earth's obliquity would be even more dramatic, also in the case of a shorter period of rotation. Had this occurred, the Earth's climate would have drastically changed on timescales of a few million years, possibly preventing the development of complex forms of life. Thus, these results seem to imply that our own existence is correlated with the existence of the Moon, which in turn is believed to be the result of a fortuitous event in the early history of the Solar System. In fact, the analogy between the Moon's composition and the Earth's mantle composition suggests that the Moon formed by accretion of material ejected from the Earth after its core–mantle differentiation (Canup and Esposito, 1996; Ida *et al.*, 1997). The ejection of a sufficient quantity of material requires the collision of the Earth with a planetary embryo of mass comparable to that of Mars. Notice also that none of the other planets (with the exception of Pluto) have a satellite with a mass ratio comparable to that of the Earth–Moon system. Celestial Mechanics meets Philosophy!

In reality, the Earth–Moon system has significantly evolved under the effect of tides over the age of the Solar System: the Moon has slowly increased the

semimajor axis of its orbit, and the spin and precession rates of the Earth have both slowed down. The dynamics of the Earth's spin axis during the tidal evolution of the Earth–Moon system has been numerically investigated by Touma and Wisdom (1994a) and by Néron de Surgy and Laskar (1997).

Laskar and Robutel (1993) also explored the dynamical evolution of the spin axes of the other planets of the Solar System, in a way analogous to that described for the Earth. For the giant planets, essentially regular solutions are obtained over a wide range of precession rates and initial obliquities. This implies in particular that Uranus' large obliquity (close to 90°) is not the result of dynamical evolution, but it is rather the consequence of a collision with an Earth-mass planetary embryo in the primordial age of the Solar System (Safronov, 1966; Parisi and Brunini, 1996). The essential regularity of the spin axes dynamics of the giant planets is due to the simple form of the spectrum of $A(t) + \iota B(t)$ for these planets, which shows a limited number of well-separated lines.

Mars' spin axis, conversely, evolves chaotically, the obliquity ranging from 0 to 60 degrees. This is due to Mars' precession frequency being close to some major secular frequencies of its orbit, as first remarked by Ward (1974). The chaoticity and the amplitude of variation of the obliquity of the Martian spin axis has been confirmed by Touma and Wisdom (1993) through the integration of the nonaveraged equations of motion.

Venus' obliquity is presently close to 180°, and its evolution is regular. Mercury is locked in the 2/3 spin orbit resonance (its orbital period is 1.5 times longer than its rotation period) and therefore the dynamics of its spin axis cannot be studied through the averaged Hamiltonian (7.16). A discussion of the past tidal evolution of the spin axes of these two planets can also be found in Laskar and Robutel (1993).

Chapter 8

SECULAR DYNAMICS OF SMALL BODIES

8.1 The linear integrable approximation

The secular normal form for the dynamics of a small body is formally equivalent to that of the planetary problem (see Section 2.5.1): it is independent of the *mean* mean longitudes of the body and of the planets, and depends only on the *mean* eccentricities, inclinations, longitudes of perihelion and node. As in the previous chapter, we will omit all the superscripts that denote that the variables appearing in the secular normal form are the *mean* modified Delaunay variables. Moreover, we will refer to the mean orbital elements defined from the mean modified Delaunay variables, omitting the adjective *mean*.

After dropping the constant term $\mathcal{H}_0(\Lambda, \Lambda_j)$ from the secular normal form, the latter can be written

$$\mathcal{H}_{\rm sec} = \varepsilon \mathcal{H}_1(P, Q, p, q, e_j, \varpi_j, i_j, \Omega_j) \tag{8.1}$$

where the terms of the normal form of higher order in ε have also been included in $\varepsilon \mathcal{H}_1$ and $e_j, \varpi_j, i_j, \Omega_j$ denote the eccentricity, the longitude of perihelion, the inclination and the longitude of node of the j-th planet.

In an attempt to give a realistic description of the secular motion of the small body, one cannot forget that $e_j, \varpi_j, i_j, \Omega_j$ slowly change with time, as shown in the previous chapter. In most applications the planets' orbital elements are assumed to evolve according to the Lagrange–Laplace solution (7.10), with the values of the coefficients listed in Tables 7.1, 7.2 and 7.3. Then, to make the Hamiltonian (8.1) autonomous, one introduces new angles $\varpi_k^* = g_k t + \beta_k$ and $\Omega_k^* = s_k t + \delta_k$. Denoting by Λ_{g_k} and Λ_{s_k} the conjugate

actions, one rewrites (8.1) as:

$$\mathcal{H}_{\rm sec} = \sum_k (g_k \Lambda_{g_k} + s_k \Lambda_{s_k}) + \varepsilon \mathcal{H}_1(P, Q, p, q, \varpi_k^*, \Omega_k^*) \ . \qquad (8.2)$$

Remember from the previous chapter that s_5 is equal to zero and that, when writing the orbital elements with respect to the invariant plane of the planetary system, $\mathcal{H}_1$ is independent of Ω_5^*.

In principle, one could assume more realistic quasi-periodic time dependencies for the planetary elements, as those given by higher-order normal forms of the secular planetary problem; this however would not increase the number of degrees of freedom, and therefore would not change the general form of (8.2), but would just modify the dependence of $\mathcal{H}_1$ on the angles ϖ_k^* and Ω_k^*. Conversely, to preserve the form (8.2) of the secular problem it is essential to consider only quasi-periodic approximations of the planetary secular motion. In the previous chapter we have seen that the secular motion of the planets is in reality not quasi-periodic; this is particularly true for the terrestrial planets. If one considered that the time dependence of the planetary orbital elements have continuous Fourier spectra, the secular problem of small-body dynamics couldn't be written as in (8.2), because it would not be possible to introduce an additional degree of freedom for each independent planetary frequency. Luckily, because the Fourier spectrum of the orbital elements of the most massive terrestrial planets is strongly peaked in correspondence to a discrete set of frequency values, the quasi-periodic approximation of planetary secular motion can still be considered as a good approximation for a realistic study of the secular dynamics of small bodies. In fact, Guzzo and Morbidelli (1997) have shown that a Nekhoroshev-like stability result for the secular motion of the small body, achieved in the framework of the quasi-periodic approximation of planetary motion, would not been altered when considering the continuous spectra of the planetary elements. For this reason, in this chapter the study of the secular small-body dynamics will be done starting from a secular normal form of type (8.2).

The Hamiltonian (8.2) is not written in quasi-integrable form, because the planetary frequencies g_k and s_k are also of order ε. Thus, as for the planetary secular problem, one needs to find an integrable approximation and to identify a new small parameter, say η, that controls the size of the perturbation. However, in this case the strategy is not unique, due to the fact that small bodies exist on a large variety of orbits, some of which have large eccentricities and/or large inclinations. Each strategy makes use of a different integrable approximation and a different small parameter. In this section we detail the "classical" integrable approximation, which is valid for those small bodies whose eccentricities and inclinations are of the same order of magnitude

as the planetary ones. In the next section an alternative strategy, suitable for the other small bodies, will be discussed.

Recalling that $P \sim \sqrt{a}e^2/2$ and $Q \sim 2\sqrt{a}\sin^2(i/2)$, where e and i denote the eccentricity and the inclination of the small body, the function $\varepsilon\mathcal{H}_1$ in (8.1) is expanded in power series of $\sqrt{P}, \sqrt{Q}$ and of $e_j, \sin(i_j/2)$. One considers $\sqrt{P}, \sqrt{Q}, e_j, \sin i_j/2$ to be of the same order of magnitude, say of order $\sqrt{\eta}$; in this way, one can write $\varepsilon\mathcal{H}_1 = \sum_n \mathcal{H}_{(n)}$, with $\mathcal{H}_{(n)}$ of order n in $\sqrt{\eta}$. Following the same reasoning as in Section 7.1, one can easily show that, because of the D'Alembert rules (see Section 1.9.3), all the terms $\mathcal{H}_{(n)}$ with odd n or $n = 0$ vanish. This implies that η is the natural "small parameter" in the resulting series expansion. Still for the D'Alembert rules, the lowest-order term of the expansion has the form:

$$\begin{aligned}\mathcal{H}_{(2)} = -g_0 P - s_0 Q + \textstyle\sum_j [\, c_j\, e_j \sqrt{2P}(\cos p \cos \varpi_j - \sin p \sin \varpi_j) \\ + d_j \sin \tfrac{i_j}{2} \sqrt{2Q}(\cos q \cos \Omega_j - \sin q \sin \Omega_j)] \ ,\end{aligned} \tag{8.3}$$

where $-g_0, -s_0, c_j, d_j$ are the coefficients of the expansion, which depend only on the semimajor axes of the small body and of the planets. The choice of denoting by $-g_0$ and $-s_0$ the coefficients of the linear terms in P and Q has been done to make the notation of the following formulæ equal to those usually found in the literature. It is curious to note that, if the secular normal form is constructed only up to order ε, s_0 turns out to be equal to $-g_0$ for all values of the semimajor axis of the small body; but, if also the terms of order ε^2 are included in the normal form, as in the work by Milani and Knežević (1990), one generally has $s_0 \neq -g_0$.

Taking into account the Lagrange–Laplace solution (7.10) for the secular evolution of the planetary elements, (8.3) becomes:

$$\begin{aligned}\mathcal{H}_{(2)} = -g_0 P - s_0 Q + \textstyle\sum_j \sum_k [\, c_j\, M_{j,k} \sqrt{2P}(\cos p \cos \varpi_k^* - \sin p \sin \varpi_k^*) \\ + d_j\, N_{j,k} \sqrt{2Q}(\cos q \cos \Omega_k^* - \sin q \sin \Omega_k^*)] \ .\end{aligned} \tag{8.4}$$

The Hamiltonian

$$\mathcal{H}_{\rm int} = \sum_k (g_k \Lambda_{g_k} + s_k \Lambda_{s_k}) + \mathcal{H}_{(2)} \tag{8.5}$$

is integrable. In fact, introducing the usual canonical polynomial variables $y = \sqrt{2P}\cos p, x = \sqrt{2P}\sin p, z = \sqrt{2Q}\cos q, v = \sqrt{2Q}\sin q$ (x and v are the new coordinates and y and z the conjugate momenta), the equations of motion

of (8.5) become

$$\begin{aligned}
\dot\Lambda_{g_k} &= \textstyle\sum_j c_j M_{j,k}[y\sin\varpi_k^* + x\cos\varpi_k^*] \ , & \dot\varpi_k^* &= g_k \\
\dot\Lambda_{s_k} &= \textstyle\sum_j d_j N_{j,k}[z\sin\Omega_k^* + v\cos\Omega_k^*] \ , & \dot\Omega_k^* &= s_k \\
\dot x &= -g_0 y + \textstyle\sum_j\sum_k c_j M_{j,k}\cos\varpi_k^* \ , & \dot y &= g_0 x + \textstyle\sum_j\sum_k c_j M_{j,k}\sin\varpi_k^* \\
\dot v &= -s_0 z + \textstyle\sum_j\sum_k d_j N_{j,k}\cos\Omega_k^* \ , & \dot z &= s_0 v + \textstyle\sum_j\sum_k d_j N_{j,k}\sin\Omega_k^* \ .
\end{aligned} \tag{8.6}$$

Using the trivial solutions $\varpi_k^* = g_k t + \beta_k$ and $\Omega_k^* = s_k t + \delta_k$, the last two lines in (8.6) become the well-known equations of two decoupled forced harmonic oscillators, whose general solution is:

$$\begin{aligned}
x &= A\sin(-g_0 t + \alpha) + \textstyle\sum_j\sum_k \xi_{j,k}\sin\varpi_k^* \\
y &= A\cos(-g_0 t + \alpha) - \textstyle\sum_j\sum_k \xi_{j,k}\cos\varpi_k^* \\
v &= B\sin(-s_0 t + \beta) + \textstyle\sum_j\sum_k \eta_{j,k}\sin\Omega_k^* \\
z &= B\cos(-s_0 t + \beta) - \textstyle\sum_j\sum_k \eta_{j,k}\cos\Omega_k^* \ ,
\end{aligned} \tag{8.7}$$

where A, α, B, β are integration constants (depending on the initial conditions) and

$$\xi_{j,k} = \frac{c_j M_{j,k}}{g_k - g_0} \ , \quad \eta_{j,k} = \frac{d_j N_{j,k}}{s_k - s_0} \ , \tag{8.8}$$

are the amplitudes of forced oscillations. Note that the latter become infinite when the frequency g_0 or s_0 is equal to one of the planetary frequencies g_k or s_k. When this happens, we say that a *resonance* occurs. Because the denominators $g_k - g_0$ and $s_k - s_0$ associated to these resonances appear in the linear equations of the small body's secular motion, these resonances are usually called *linear secular resonances* or *secular resonances of first order.* It is also quite common to denote by ν_k the resonance $g_0 = g_k$ and by ν_{1k} the resonance $s_0 = s_k$. The reader might be surprised that the amplitudes (8.8) become infinite corresponding to these resonances, because in Chapter 4 it was shown that resonances give pendulum-like dynamics, with bounded oscillation amplitudes. The divergence of the forced oscillation amplitudes (8.8) is here an artifact due to the *linearity* of the equations of motion (8.6) with respect to the variables x, y, v, z. Remember that these equations are derived from the Hamiltonian (8.5), where the quadratic terms in the actions P and Q have not been retained. Similarly, if the torque $\beta/2\hat I_1^2$ had been neglected in Hamiltonian (4.4), the motion of $\hat I_1$ would have been unbounded. The accurate study of the dynamics in secular resonances, which takes into account also the higher-order terms in the action variables, will be done in Section 8.4.

We now take advantage of the solution (8.7) to transform (8.5) into a Hamiltonian that depends only on action variables, under the hypothesis that no linear secular resonance occurs. We first look for new canonical variables $y', x', z', v', \Lambda'_{g_k}, {\varpi^*}'_k, \Lambda'_{s_k}, {\Omega^*}'_k$ so that (8.5) becomes

$$\mathcal{H}_{\rm int} = \sum_k (g_k \Lambda'_{g_k} + s_k \Lambda'_{s_k}) - g_0 \frac{x'^2 + y'^2}{2} - s_0 \frac{v'^2 + z'^2}{2} + F({\varpi^*}'_k, {\Omega^*}'_k) \ . \quad (8.9)$$

Comparing the time evolution of $x, y, v, z, \varpi^*_k, \Omega^*_k$ given by (8.7) with that of $x', y', v', z', {\varpi^*}'_k, {\Omega^*}'_k$ given by the Hamiltonian equations of (8.9), it is natural to choose

$$\begin{aligned} x' &= x - \textstyle\sum_j \sum_k \xi_{j,k} \sin \varpi^*_k \ , \quad & y' &= y + \textstyle\sum_j \sum_k \xi_{j,k} \cos \varpi^*_k \\ v' &= v - \textstyle\sum_j \sum_k \eta_{j,k} \sin \Omega^*_k \ , \quad & z' &= z + \textstyle\sum_j \sum_k \eta_{j,k} \cos \Omega^*_k \end{aligned} \quad (8.10)$$

together with ${\varpi^*}'_k = \varpi^*_k$, ${\Omega^*}'_k = \Omega^*_k$. Then, we just have to define Λ'_{g_k} and Λ'_{s_k} in order to have a canonical transformation. For this purpose, we use the generating function criterion (see Section 1.6, and equation (1.41) in particular): the transformation of the variables reported above can be rewritten in the form of implicit equations $x' = \partial S/\partial y', y = \partial S/\partial x, \ldots, {\Omega^*}'_k = \partial S/\partial \Lambda'_{s_k}$, if one chooses

$$\begin{aligned} S(x, y', v, z', \varpi^*_k, \Lambda'_{g_k}, \Omega^*_k, \Lambda'_{s_k}) = \textstyle\sum_k (\varpi^*_k \Lambda'_{g_k} + \Omega^*_k \Lambda'_{s_k}) + xy' + vz' \\ - \textstyle\sum_j \sum_k [\xi_{j,k}(y' \sin \varpi^*_k + x \cos \varpi^*_k) \\ + \eta_{j,k}(z' \sin \Omega^*_k + v \cos \Omega^*_k)] \ . \end{aligned} \quad (8.11)$$

This proves that the transformation is canonical, provided the transformation from $\Lambda_{g_k}, \Lambda_{s_k}$ to $\Lambda'_{g_k}, \Lambda'_{s_k}$ is given by the equations $\Lambda_{g_k} = \partial S/\partial \varpi^*_k$ and $\Lambda_{s_k} = \partial S/\partial \Omega^*_k$. By construction, the Hamiltonian (8.5) is transformed into the form (8.9); the function F will be dropped from the Hamiltonian, because it does not depend of the small body's variables x', y', v', z'.

It is now sufficient to introduce new canonical action–angle variables P', p', Q', q' through the equations

$$x' = \sqrt{2P'} \sin p' \ , \ y' = \sqrt{2P'} \cos p' \ , \ v' = \sqrt{2Q'} \sin q' \ , \ z' = \sqrt{2Q'} \cos q' \quad (8.12)$$

to transform (8.9) into

$$\mathcal{H}_{\rm int} = \sum_k (g_k \Lambda'_{g_k} + s_k \Lambda'_{s_k}) - g_0 P' - s_0 Q' \ , \quad (8.13)$$

i.e. into a Hamiltonian that depends only on the action variables, which is ready to be used as an integrable approximation for the perturbative study of the full secular normal form. The new action–angle variables are usually called *linear proper modified Delaunay variables.* The elements introduced by inversion of (1.69) and (1.68), with P', p', Q', q' instead of P, p, Q, q, namely:

$$\begin{aligned} e' &= \sqrt{1 - \left[1 - \tfrac{P'}{L}\right]^2} \ , & \varpi' &= -p' \\ i' &= \text{arc cos} \left[1 - \tfrac{Q'}{L-P'}\right] \ , & \Omega' &= -q' \ , \end{aligned} \tag{8.14}$$

are respectively called the *linear proper* eccentricity, perihelion longitude, inclination and node longitude. From (8.13) it is evident that g_0 and s_0 are the frequencies of the linear proper perihelion and node longitudes respectively; they are therefore called the *linear proper frequencies.*

8.2 The Kozai integrable approximation

We now look for an integrable approximation of (8.2) which is suitable for small bodies on orbits with eccentricity and/or inclination that are much larger than those of the planets. In this case, the actions P and/or Q introduced in (1.69) are not small, so that it is not reasonable to expand the Hamiltonian in a power series of $\sqrt{2P}$ and $\sqrt{2Q}$ as in the previous section. Instead, following a scheme originally developed by Williams (1969), we expand (8.2) in power series of only e_j and $\sin i_j/2$, writing

$$\mathcal{H}_{\rm sec} = \sum_k (g_k \Lambda_{g_k} + s_k \Lambda_{s_k}) + \sum_{n\geq 0} \mathcal{K}_{(n)}(P, Q, p, q, \varpi_k^*, \Omega_k^*) \ , \tag{8.15}$$

with $\mathcal{K}_{(n)}$ of degree n in $e_j, \sin i_j/2$. The leading term of this expansion is $\mathcal{K}_{(0)}$. In this case the "small parameter" η that characterizes the size of the perturbation relative to the main part of the Hamiltonian is of the order of the planetary eccentricities and inclinations (remember that it is of the order of e_j^2, i_j^2 in the approach based on the linear approximation).

The zero-order term $\mathcal{K}_{(0)}$ is independent of the planetary eccentricities and inclinations, so that – for the D'Alembert rules – it is also independent of the planetary angles ϖ_k^* and Ω_k^*. The D'Alembert rule on the rotational invariance of the Hamiltonian (rule 2 in Section 1.9.3) implies that in the Fourier expansion of $\mathcal{K}_{(0)}$ only the harmonics $\exp \iota[mp + kq]$ with $m + k = 0$ may have coefficients that are not null; this implies that $\mathcal{K}_{(0)}$ depends only on the argument of perihelion $g = q - p$. Depending only on one angle, $\mathcal{K}_{(0)}$

is therefore integrable (see Section 1.8). It is then suitable to reintroduce the Delaunay variables G, H, g, h (see 1.68) and to identify the Hamiltonian

$$\mathcal{H}_{\text{int}} = \sum_k (g_k \Lambda_{g_k} + s_k \Lambda_{s_k}) + \mathcal{K}_{(0)}(G, H, g) \tag{8.16}$$

as the integrable approximation of the secular Hamiltonian (8.2). To write (8.16) as a function of the sole action variables, as needed in order to study the secular normal form using the tools of Chapter 2, it is necessary to introduce new action–angle variables following the recipe of Arnold (see Section 1.9). For this step, which will be discussed in Section 8.2.3, it is necessary to first study in detail the dynamics described by the Hamiltonian $\mathcal{K}_{(0)}$. This was done for the first time by Kozai (1962, 1978), so that $\mathcal{K}_{(0)}$ is now usually called the *Kozai Hamiltonian*.

Because $\mathcal{K}_{(0)}$ is independent of h, the action $H = \sqrt{a(1-e^2)}\cos i$ is a constant of motion. Remember that H is the z component of the small body's angular momentum. Going back to the mean elements of the small body a, e and i, and recalling that a is constant, this implies that the evolution of e and i, which are functions of the argument of perihelion g, are coupled through the relationship $H =$constant. The natural coordinates on a level surface of H are G and g, or, equivalently, e and g. On such a surface, the evolution of e (or G) as a function of g follows a level curve of $\mathcal{K}_{(0)}(G, H, g)$, where H plays the role of a parameter (see Section 1.8).

If the normal form is computed only up to order ε and the terms of higher order in ε are neglected, the practical recipe for the computation of the level curves of $\mathcal{K}_{(0)}$ is very simple. In fact, in this case the normalized Hamiltonian is just the average of the perturbation in (1.29) over the mean anomalies l and l_j of the small body and of the planets, namely

$$\mathcal{H}_{\text{sec}} = -\frac{\mathcal{G}}{(2\pi)^2} \sum_{j=1}^{N} m_j \int_0^{2\pi} \int_0^{2\pi} \left(\frac{1}{\|\boldsymbol{\Delta}_j\|} - \frac{\mathbf{r} \cdot \mathbf{s}_j}{\|\mathbf{s}_j\|^3} \right) \mathrm{d}l \, \mathrm{d}l_j \ , \tag{8.17}$$

where the vectors $\mathbf{r}$ and $\mathbf{s}_j$ denote respectively the heliocentric positions of the small body and of the j-th planet, and $\boldsymbol{\Delta}_j = \mathbf{r} - \mathbf{s}_j$; these vectors must be expressed as functions of the orbital elements. The Kozai Hamiltonian $\mathcal{K}_{(0)}$ is therefore given by (8.17), where the eccentricities and the inclinations of the planets have been preliminarily set equal to 0. Under the assumption $e_j = i_j = 0$, we know from the argument discussed above that the result of the double integral in (8.17) is independent of h, and therefore depends only on a, e, i, g and a_j. The values of a and a_j are fixed, because the semimajor axes are constant for the secular Hamiltonian. Moreover, the condition of being on the selected surface $H =$constant sets $i = \arccos(H/\sqrt{a(1-e^2)})$.

As a consequence, the value of $\mathcal{K}_{(0)}$ is completely determined once e and g are given. From the technical point of view, the integral over l_j in (8.17) can be analytically computed using the elliptic function K of the first kind. Indeed, setting the eccentricity e_j and the inclination i_j of the planet equal to zero, one has (Williams, 1969, Bailey *et al.*, 1992):

$$\int_0^{2\pi}\left(\frac{1}{\|\boldsymbol{\Delta}_j\|}-\frac{\mathbf{r}\cdot\mathbf{s}_j}{\|\mathbf{s}_j\|^3}\right)\mathrm{d}l_j=\int_0^{2\pi}\frac{1}{\|\boldsymbol{\Delta}_j\|}\mathrm{d}l_j=\frac{4}{\sqrt{r^2+a_j^2}}\sqrt{1-\frac{m}{2}}K(m)\ , \tag{8.18}$$

where

$$m=\frac{4a_j\sqrt{x^2+y^2}}{r^2+a_j^2+2a_j\sqrt{x^2+y^2}} \tag{8.19}$$

$\mathbf{r}$ is the position vector of the small body, x and y are the coordinates of its projections on the plane of the planetary orbit, and $r=\|\mathbf{r}\|$. The integral over l of (8.17), on the contrary, cannot be analytically computed, but requires some numerical algorithm (see for instance Piessens *et al.*, 1983). It is simple to write computer code that calculates $\mathcal{K}_{(0)}$ at every point e, g of a grid on the surface $H=$constant, and then to plot the level curves of $\mathcal{K}_{(0)}$ with some graphics software.

An analytic approximation of the Kozai Hamiltonian has been given by Kinoshita and Nakai (1999):

$$\mathcal{K}_{(0)}=\sum_j\frac{m_jn_j^2}{16(m_j+m_c)}a^2\left((2+3e^2)(3\cos^2 i-1)+15e^2\sin^2 i\cos 2g\right)\ , \tag{8.20}$$

where m_c is the mass of the central body and n_j is the mean motion of the j-th perturber. This expression is valid for every value of e and i, but neglects terms of order $(a/a_j)^2$, which severely limits the possibilities of applications. The expression (8.20) is useful only to study satellite problems, where the ratio a/a_j is very small.

To analyze the dynamics given by the Kozai Hamiltonian it is convenient to separate the study into two cases: the asteroidal case, where the semimajor axis of the small body's orbit is smaller than that of the main perturber, and the cometary case, which corresponds to the opposite situation.

8.2.1 Kozai dynamics inside the orbit of the main perturber

Figure 8.1 shows the evolution of e and g given by the Kozai Hamiltonian, on six surfaces $H=$constant. The dynamics have been computed for $a=3$ AU, taking into account the perturbations given by the four giant planets, so that it

well represents a typical asteroidal case. Each surface $H =$constant in Fig. 8.1 is identified by the (maximal) value of the inclination, corresponding to $e = 0$, namely $i_{\max} = \arccos(H/\sqrt{a})$. The bounding circle on each panel corresponds to the maximal value allowed for the eccentricity on the surface $H =$constant, i.e. $e_{\max} = \sqrt{1 - H^2/a}$.

Note that, as a general rule, all figures are symmetric with respect to the axis $\sin g = 0$. This is because the expansion of $\mathcal{K}_{(0)}$ in a power series of the small body's inclination presents, for the D'Alembert rules, only terms i^n with even n; therefore, the Fourier expansion of $\mathcal{K}_{(0)}$ must present only harmonics $\cos kg$ with even k. This implies that $\mathcal{K}_{(0)}$ is periodic with period π in g.

For a small value of $i_{\max}$ (Fig. 8.1a), the level curves of $\mathcal{K}_0$ are very close to circles. This means that the argument of perihelion circulates over 360 degrees, while the eccentricity (and, by consequence, also the inclination) is almost constant. This shows that the linear integrable approximation of the secular Hamiltonian (see Section 8.1), which neglects all harmonics in the perihelion argument, is in this low-inclination case a good approximation of Kozai's dynamics.

Increasing the value of $i_{\max}$ (Fig. 8.1b), the level curves elongate along the $e \sin g$ axis. The argument of perihelion still circulates, but the eccentricity oscillates, assuming the maximum value each time that $g = 90$ or 270 degrees; correspondingly, the inclination assumes a minimum value. Above a critical threshold of $i_{\max}$ the structure of Kozai's dynamics abruptly changes (Fig. 8.1c). The point $e = 0$ becomes an unstable equilibrium point; a separatrix connects this point to itself and divides the phase space into three parts: two regions where g librates around either 90 or 270 degrees (i.e. its precession stops) and one region where g circulates. This dynamical structure is typical of the integrable resonant case (Chapter 4), so that it is usually referred to as the *Kozai resonance*. The Kozai resonance can be regarded as a 1:1 resonance between the precession frequencies of the longitude of perihelion ϖ and of the longitude of node Ω of the small body. Therefore, in the notation of Chapter 4, the argument of perihelion $g \equiv \varpi - \Omega$ is the critical angle of the resonance. It is evident that, when the Kozai resonance appears, the classical integrable approximation (Section 8.1) is no longer a good approximation of the real secular dynamics.

Figures 8.1d, 8.1e and 8.1f correspond to larger values of $i_{\max}$. The Kozai resonance becomes stronger, in the sense that the width of the libration regions increases. The point at $e = 0$ is always an unstable equilibrium point. As a consequence, any orbit with initial small eccentricity is forced by the resonant dynamics to reach a large value of e when $g = 90^\circ$ or 270°. Large eccentricity implies small heliocentric nodal distance, the latter being the distance from

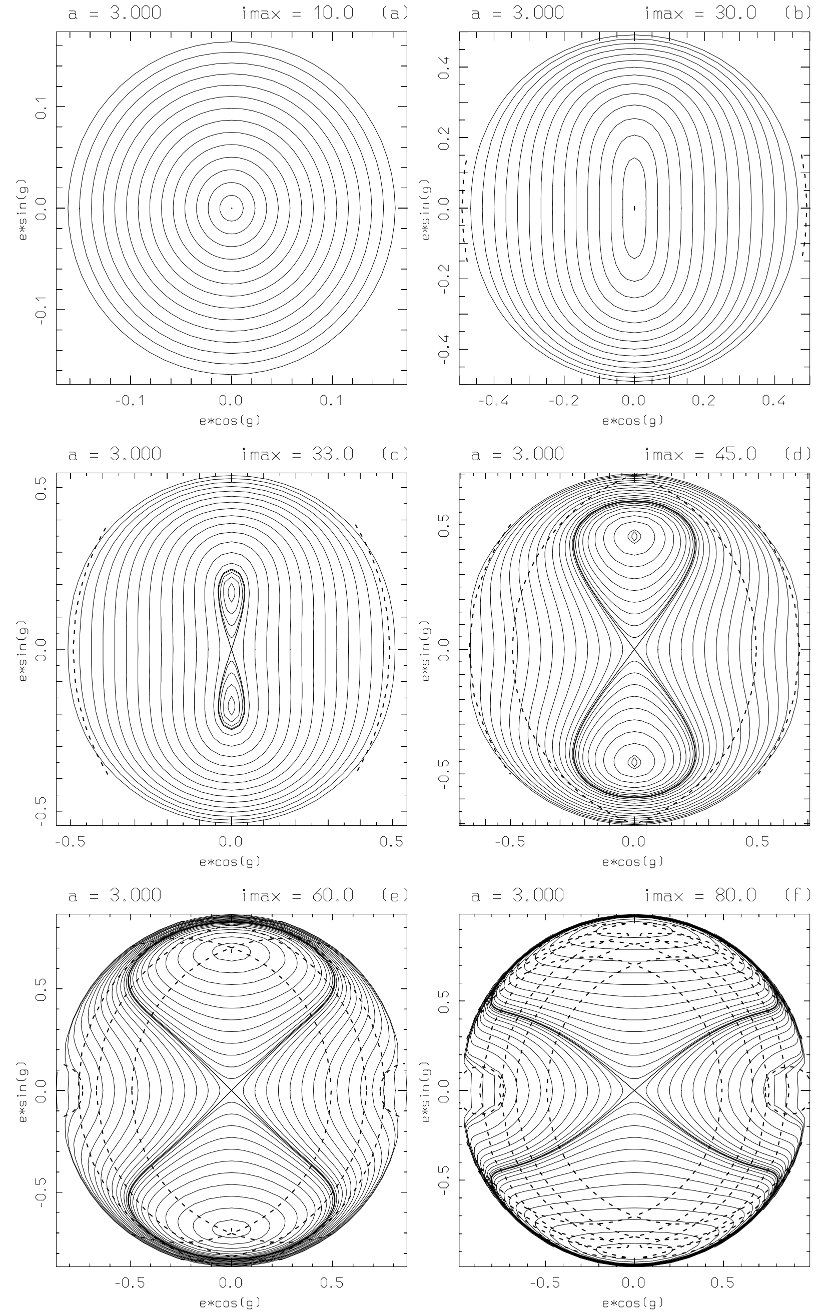


Figure 8.1: The Kozai dynamics in the asteroid belt, represented in polar coordinates e, g. All panels are plotted for a value of the mean semimajor axis equal to 3 AU. The inclination $i_{\max}$ corresponding to $e = 0$ is shown on the top of each panel. The dashed curves denote the values of e, g corresponding to nodal crossing of the orbits of the planets. Adapted from Thomas (1998).

the Sun when crossing the reference plane, as given by the formula

$$r_{\pm} = \frac{a(1-e^2)}{1 \pm e \cos g} \tag{8.21}$$

where + and − refer respectively to the distance at the ascending and descending node. Therefore, the asteroid must cross, at some point of its secular evolution, the orbit of the terrestrial planets. In Fig. 8.1c, the dashed curves denote the set of points e, g that give a node-crossing of the orbit of Mars (simply plotted by setting $r_{\pm} = 1.5$AU in (8.21)); in Fig. 8.1d the corresponding curves for node-crossing of the orbits of the Earth are also plotted; the curves denoting node-crossing of Venus and Jupiter are introduced in Fig. 8.1e and those related to Mercury are visible in Fig. 8.1f. A node-crossing of the orbit of a planet opens the possibility for close encounters between the planet and the asteroid, which destabilize the orbit of the latter. The Kozai dynamics therefore explains why the asteroid belt is depleted at large inclination. From the mathematical viewpoint, the close encounter corresponds to a true singularity ($\boldsymbol{\Delta}_j = 0$) of the Hamiltonian (1.29). The analytic construction of the secular normal form therefore breaks down. It is however possible to show (Gronchi and Milani, 1998, 1999) that the double integrals defining the Kozai Hamiltonian in (8.17) are well defined even in the presence of such a singularity. Therefore, the Kozai Hamiltonian can still be used for a first insight into the secular dynamics of planet-crossing orbits (as will be better seen in Section 8.2.2).

The value of the critical threshold $i_{\max}$ for the appearance of the Kozai resonance depends on a, and its precise determination can be found in Kozai (1962). The Kozai dynamics turns out to be very similar to that shown in Fig. 8.1 for all ratios a/a' between 0 and 1, where a denotes the semimajor axis of the small body and a' that of the main perturber (Jupiter in the asteroid belt case). The dynamics described above is therefore not specific to the asteroid belt, but concerns a large variety of systems. For instance, the Soviet artificial satellite *Lunik*, initially put on a quasi-polar orbit around the Earth, crashed to the ground after a few days because the Kozai resonance generated by the joint Lunar and Solar perturbations forced its eccentricity to increase so much that the perigee became smaller than the Earth's radius. In principle, in any dynamical system that can be described by a restricted three-body problem, the Kozai resonance causes the instability of the inner body if its inclination with respect to the outer perturber is large enough. This is true despite the distance or the mass of the outer perturber: indeed, as an effect of the degeneracy of the two-body problem, the strength of the perturbation (ε in formula 8.1) just reflects on the timescale of Kozai's dynamics, but not on its effects. However, if other perturbations force the precession of the

argument of perihelion g, then the the secular Hamiltonian becomes, to a first approximation,

$$\mathcal{H}_{\rm sec} = \alpha G + \mathcal{K}_{(0)}(G, H, g) \tag{8.22}$$

where the term α is the forced precession frequency of g given by the additional perturbations and $\mathcal{K}_{(0)}$ is the Kozai Hamiltonian due to the distant perturber. If $\mathcal{K}_{(0)}$ is small with respect to α, then the Kozai resonance disappears, whatever the inclination is. An example of this is provided by the satellites of Uranus. Their orbits being close to the equatorial plane of the planet, these bodies have an inclination of approximately 90 degrees with respect to the plane of the Sun's apparent motion. Therefore, the Kozai resonance generated by the solar perturbation should force them to fall onto the planet. However, the precession frequencies of their arguments of pericenter, forced by the oblateness of the planet and by their mutual perturbations, are large enough to suppress the Kozai resonance, thus allowing the long-term orbital stability of the satellites. The same happens for the entire planetary system. The planets' orbits are in fact inclined by about 60 degrees with respect to the Galactic plane. If the Solar System had only one planet, the perturbation given by the over-density of matter on the Galactic plane (the so-called *Galactic tide*, see for instance Matese and Whitman, 1992) would periodically pump the planet's eccentricity. However, the mutual perturbations among the planets force – as seen in the previous chapter – nonnegligible precession frequencies of the arguments of perihelion, so that the effect of the Galactic tide becomes negligible. Conversely, for a body with semimajor axis equal to a few thousand astronomical units, the precession frequency forced by the planets becomes very small, and the Galactic tide becomes the dominating gravitational perturbation. In fact, the most credited theory for the origin of the Oort cloud (the reservoir of long-period comets) invokes Kozai-like dynamics generated by the Galactic tide as responsible for decreasing the eccentricities of the bodies that have been scattered by the planets onto orbits with semimajor axes equal to several 10^3 AU (see Duncan *et al.*, 1987).

8.2.2 Kozai dynamics outside the orbit of the main perturber

Figure 8.2 is the analog of Fig. 8.1, but has been computed for a small body with semimajor axis equal to 45 AU, namely exterior to its perturbers (the four giant planets). At a glance, it is manifest that in the outer Solar System the Kozai dynamics is very different in comparison with that of the asteroid belt. When $i_{\rm max}$ is small the argument of perihelion always circulates (Fig. 8.2a) and the eccentricity does not change in a significant way (the level curves of $\mathcal{K}_{(0)}$ are almost circles), similarly to the asteroid case. However, increasing $i_{\rm max}$

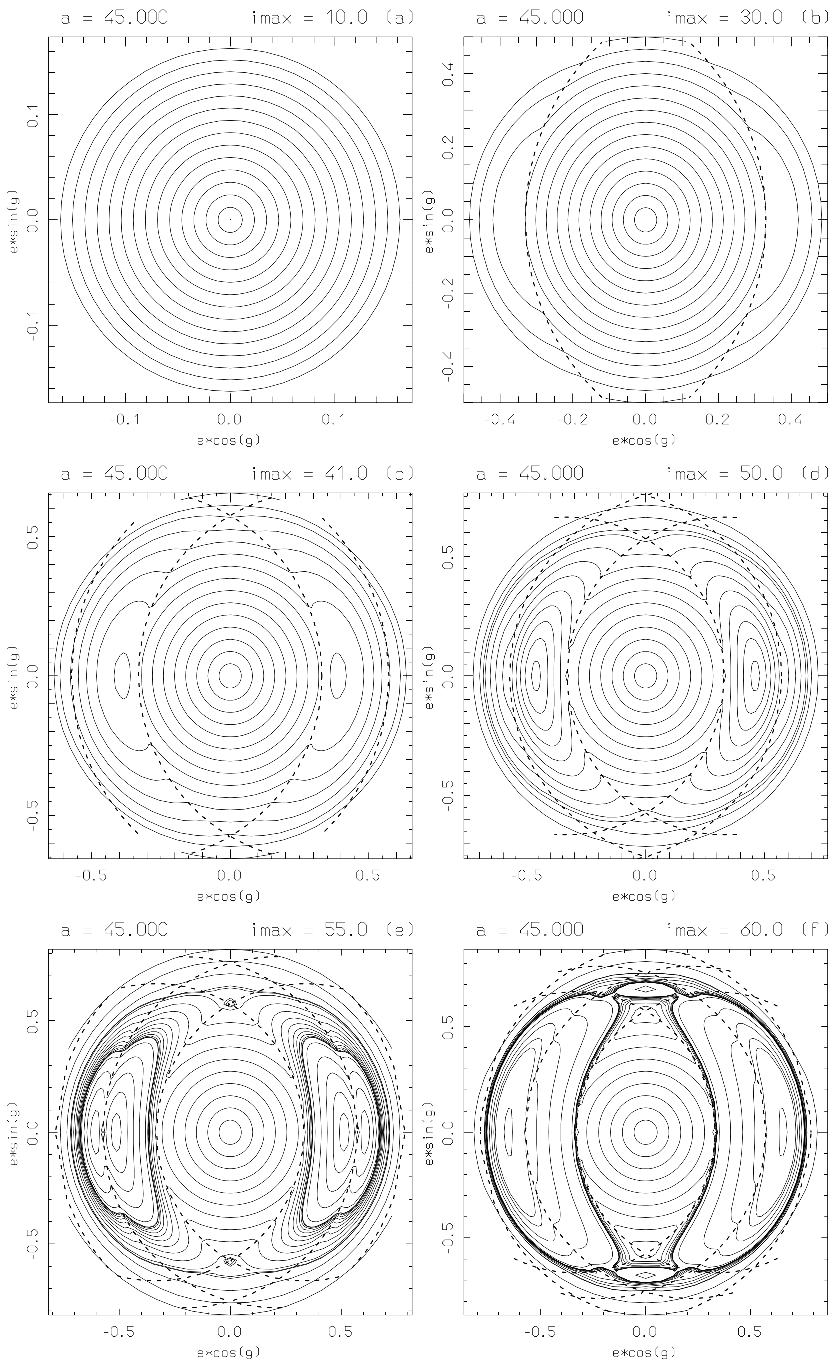


Figure 8.2: The same as Fig. 8.1 but for a semimajor axis typical of the Kuiper belt. The dashed curves denote the values of e, g corresponding to node-crossing of the outer planets' orbits. Adapted from Morbidelli and Thomas (1996).

(following panels in Fig. 8.2) at low eccentricities the level curves persist to be quasi-circular. In Fig. 8.2b the dashed curves report the families of values (e, g) corresponding to node-crossing of the orbit of Neptune. The low-eccentricity orbits are always protected from encounters with Neptune; conversely, the circulation of g takes all high-eccentricity orbits to cross the orbit of the planet. Note that the level curves of $\mathcal{K}_{(0)}$ have angular points in correspondence with the Neptune crossing line: this is because the double averaged function (8.17) is continuous for $\mathbf{\Delta}_j = 0$, but not differentiable (Gronchi and Milani, 1998, 1999).

On the surface $H =$ constant corresponding to $i_{\max} = 40°$ (Fig. 8.2c) the phase space changes, because a couple of stable equilibrium points are formed at $g = 0°$ and $g = 180°$. These equilibria are surrounded by islands where the argument of perihelion librates. Among all high-eccentricity orbits, only those in these two islands of libration are protected from encounters with Neptune. Encounters with Uranus are also possible for very high-eccentricity orbits (the new dashed curves denote the node-crossing of Uranus' orbit). In Fig. 8.2d, the islands of libration are much wider, and orbits with large amplitude of libration are forced to cross the orbit of Uranus. In Fig. 8.2e the curves denoting Uranus-crossing orbits cut each island of libration in two parts, creating four small islands of libration protected from planetary encounters. At very large eccentricity, also the dashed curves corresponding to encounters with Saturn appear. Finally, in Fig. 8.2f only two islands of libration at $g = 0°, 180°$ survive beyond the Uranus-crossing curve. Moreover we note the formation of two additional islands of librations at $g = 90°, 270°$, bounded by the curves denoting node-crossing of the orbits of Neptune and Uranus.

All the panels shown in Fig. 8.2 allow the conclusion that, in contrast to the asteroid belt case, an initially small to moderate eccentricity would not be increased under the effect of the Kozai resonance, whatever the inclination is. Thus, small-eccentricity orbits stay protected from planetary encounters. In fact, in the Kuiper belt there seem to be several bodies with inclinations much larger than in the asteroid belt. Conversely, the Kozai dynamics is nontrivial at large eccentricity, providing islands of stable motion protected from planetary encounters. It is therefore primarily interesting for the secular evolution of long-period comets. To cover the range of eccentricities typical of cometary orbits, one has to focus on the surfaces $H =$ constant corresponding to $i_{\max} > 60°$. Figure 8.3 presents the level curves of $\mathcal{K}_{(0)}$ on these surfaces using Cartesian coordinates g and $X = \sqrt{1 - e^2}$. The advantage of these coordinates is to magnify the high-eccentricity region with respect to the low-eccentricity region, X being a highly nonlinear function of e.

Figure 8.3a corresponds to the same surface illustrated in Fig. 8.2f, so

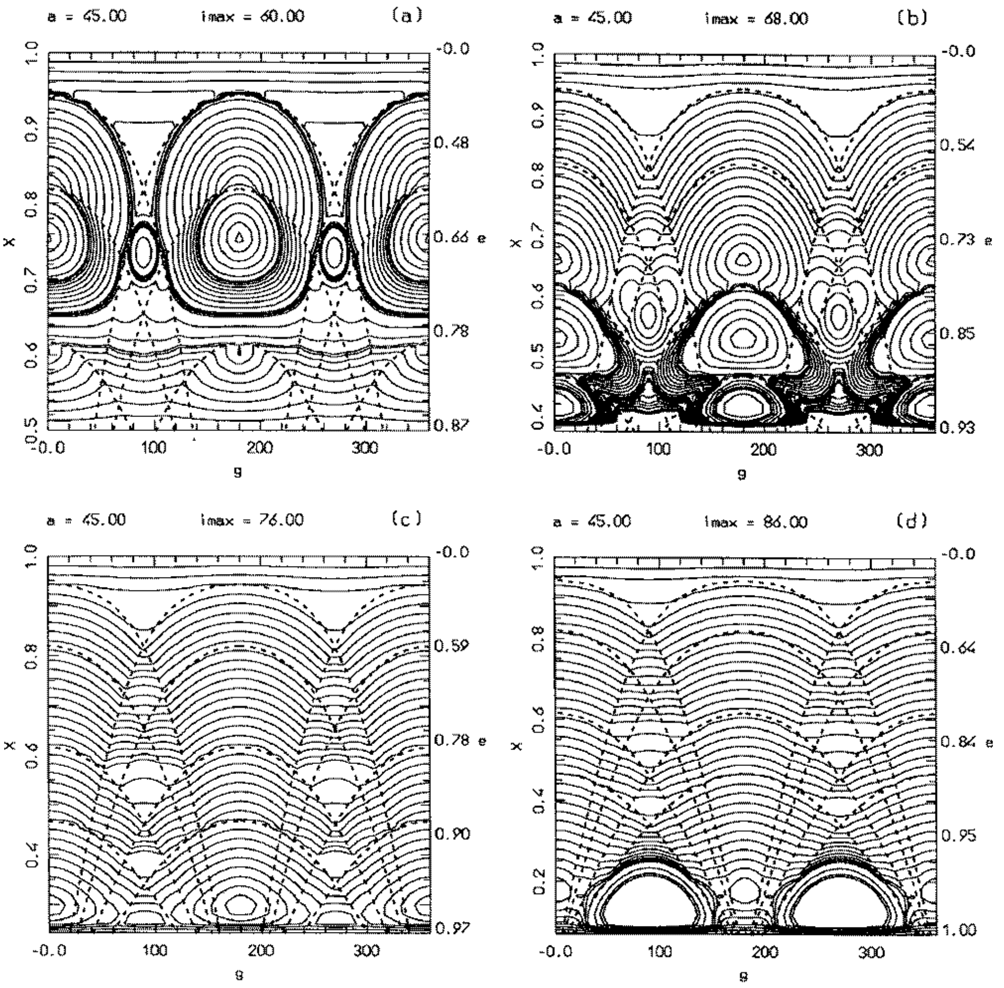


Figure 8.3: Kozai dynamics in the Kuiper belt, for very large eccentricities and inclinations. The y scale is linear in $X = \sqrt{1 - e^2}$ to magnify the large eccentricity region of the phase space. Panel (a) is plotted for the same value of $i_{\max}$ of panel (f) in Fig. 8.2. Reprinted from Fig. 5 of Thomas and Morbidelli (1996), with permission from Kluwer Academic Publishers.

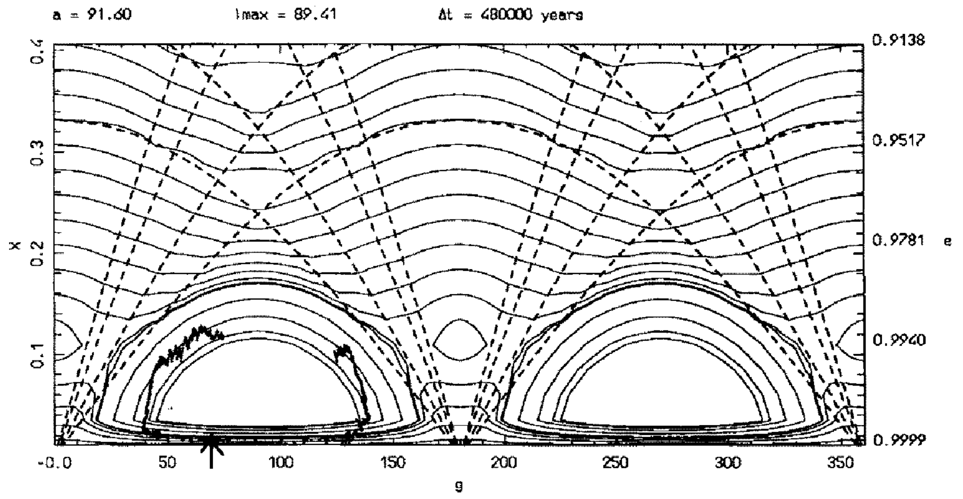


Figure 8.4: The Kozai dynamics computed for the values of the semimajor axis and $i_{\rm max}$ of the Sun-grazing comet Ikeya–Seki. Superposed is the evolution of the comet computed by numerically integrating the full equations of motion, over a timespan of 480,000 y. Reprinted from Fig. 8 of Thomas and Morbidelli (1996), with permission from Kluwer Academic Publishers.

that the reader can establish a direct correspondence between the two coordinate systems. In Fig. 8.3b also the curves corresponding to node-crossing of the orbit of Jupiter are marked. Each of the big islands of libration at $g = 0°, 180°$ is broken into three islands by the curves of node-crossing of Saturn's and Jupiter's orbits. Each island is protected from planetary encounters. In Fig. 8.3c, the two upper islands at $g = 0°(180°)$ are destroyed and only the third island below the Jupiter-crossing curves survives. But the curves denoting node-crossing of the orbits of Neptune and Uranus become closer, so that this surviving island is now smaller than in panel b. Conversely a new island forms, still below the Jupiter-crossing curves but with $g = 90°$ $(270°)$, and it becomes larger with increasing value of $i_{\rm max}$. Indeed in Fig. 8.3d, corresponding to $i_{\rm max} = 86°$, the island at $g = 90°$ $(270°)$ dominates the phase space, while that at $g = 0°$ $(180°)$ almost disappears.

The features of the Kozai dynamics illustrated in Figs. 8.2 and 8.3 are typical of the outer Solar System and do not change qualitatively with the semimajor axis of the small body. This is quite intuitive. The appearance and disappearance of the islands is dictated by the position of the curves denoting node-crossing of planetary orbits. Thus, increasing the semimajor

axis, the same situation simply repeats at larger eccentricity and at the same inclination, i.e. for a larger value of $i_{\max}$.

As first pointed out by Bailey *et al.* (1992), the libration islands at $g = 90°, 270°$ visible in Fig. 8.3d could be responsible for the origin of many Sun-grazing comets, as well as of the comets that collide with the Sun like those that are frequently discovered by the Soho satellite. The latter are long-period comets with extremely large eccentricities, so that their perihelion distances are of the order of the Sun's radius. They are expected to have reached their current eccentricities by following half a libration cycle in one of the two main islands of Fig. 8.3d. As an example, Fig. 8.4 shows the case of the Sun-grazer 1965VIII Ikeya–Seki, one of the most magnificent comets of the XXth century: the panel shows the Kozai dynamics corresponding to the present orbital elements of the comet ($a = 91.6$ AU, $e = 0.9999$, $i = 141.86°$, $g = 69.05°$) and the arrow denotes the position of the comet in g, X coordinates; superposed, the past and future evolutions of the comet are shown, as resulting from an accurate numerical integration of the full equations of motion covering 480,000 y. As one sees, the numerically computed evolution basically follows the level curves of the Kozai Hamiltonian; the agreement is not perfect because the first-order secular normal form is a rough approximation of the real motion, these comets being strongly perturbed when passing through the planetary system; moreover the Kozai Hamiltonian misses the secular dynamical effects due to planetary eccentricities and inclinations. A detailed discussion of the pertinence of the Kozai Hamiltonian to describe the long-term evolution of long-period comets can be found in Thomas and Morbidelli (1996); nevertheless, Fig. 8.4 shows that the simple analytic computation of the Kozai's dynamics is enough to understand the origin of the comet's Sun-grazing state.

Michel and Thomas (1996) have shown that a Kozai dynamics similar to that shown in Fig. 8.2c characterizes the motion of the asteroids with moderate inclination, and semimajor axes slightly larger than those of the Earth or Venus. The libration islands around $g = 0°, 180°$ are important because they (temporarily) protect these asteroids from close encounters with the terrestrial planet. One should not be surprised that in these cases the Kozai dynamics looks like that discussed in this section: in fact, for these values of semimajor axes and inclinations, the Earth (or Venus) becomes the main perturber, predominating over Jupiter's gravitational field. Only at large inclination do Jupiter's gravitational perturbations become predominant, and the Kozai's dynamics looks again as in Fig. 8.1d.

The Kozai Hamiltonian for the Centaurs objects, which have semimajor axes in the outer planetary region (5–30 AU) has been studied in detail

by Gronchi and Milani (1999); as for the long-period comets, the dynamics presents several libration islands, determined by the planet-crossing curves.

8.2.3 Action–angle variables for the Kozai Hamiltonian

The Kozai Hamiltonian being integrable, it is possible to introduce suitable Arnold action–angle variables in order to transform $\mathcal{K}_{(0)}$ into a Hamiltonian depending only on the new actions. As discussed in Section 4.2, different sets of action–angle variables need to be introduced in the regions where g circulates or librates, the transformation to the new variables being singular on the separatrix of the Kozai resonance.

Because the action H is a constant of motion for the Kozai Hamiltonian, one of the two Arnold actions, say J_2, is H itself. For the computation of the other action, say J_1, the cycles used in the Arnold–Liouville theorem are those identified by the level curves of $\mathcal{K}_{(0)}$. Therefore one has:

$$J_1 = \frac{1}{2\pi}\int_0^{2\pi} G(\mathcal{K}_{(0)}, H, g)\mathrm{d}g \ , \tag{8.23}$$

for a circulation cycle and

$$J_1 = \frac{1}{2\pi}\left[\int_{g_{\min}}^{g_{\max}} G^+(\mathcal{K}_{(0)}, H, g)\mathrm{d}g - \int_{g_{\min}}^{g_{\max}} G^-(\mathcal{K}_{(0)}, H, g)\mathrm{d}g\right] \ , \tag{8.24}$$

for a libration cycle. In the above formulæ $G(\mathcal{K}_{(0)}, H, g)$ denotes the dependence of G on g, along the cycle characterized by the values of the constants H and $\mathcal{K}_{(0)}$; G^+ and G^- denote respectively the upper and lower part of the cycle in the case of libration, and $g_{\min}, g_{\max}$ are the extreme values assumed by g during the libration.

Concerning the new angles ψ_1 and ψ_2, canonically conjugate to J_1 and J_2, Arnold's theory shows that they are linear functions of time, whose time derivatives are equal respectively to the frequency of circulation/libration of g and to the frequency of h averaged over a period of g. The introduction of the new action–angle variables in the full secular normal form is done as explained in Section 4.2; the integrable approximation (8.16) becomes

$$\mathcal{H}_{\text{int}} = \sum_k (g_k \Lambda_{g_k} + s_k \Lambda_{s_k}) + \mathcal{K}_{(0)}(J_1, J_2) \ . \tag{8.25}$$

For the cycles which intersect a planet-crossing curve, the introduction of the new action–angle variables can still be done in principle. In fact, the cycle is (artificially) well defined, so that the integral defining J_1 can be computed; the frequencies of the angles g and h are infinite on the planet-crossing

curves, but their mean values are still finite, when averaged over the cycle (Gronchi and Milani, 1998, 1999). However, the Hamiltonian written in the new variables is not analytic for all the values of J_1, J_2 that correspond to planet-crossing cycles. In fact, in reality the body will have a strongly chaotic motion, due to close encounters with the planet.

8.3 Proper elements

As discussed in the Introduction, the usual plot of the distribution of the asteroids with respect to their osculating semimajor axis, eccentricity and inclination (Fig. 2) shows that some regions of the belt are more densely populated than others. However, because the osculating elements continuously change under the perturbations of the planets, the mutual distances of the asteroids in the space of osculating a, e, i change on the timescale of the orbital precessions; it is therefore difficult to conclude anything about the significance of these regions of apparent overdensity. However, for the asteroids which evolve on regular orbits, i.e. on KAM tori, it is possible in principle to introduce action–angle variables such that the actions are constants of motion. In the space of these actions, the distribution of the asteroids would be invariant with time, so that its analysis could provide important information on the structure of the asteroid belt (see Section 8.3.1).

The computation of the action variables that are constant on KAM tori is beyond all practical possibilities, but fortunately it is possible to compute – through the construction of low-order Birkhoff normal forms – actions whose amplitudes of temporal oscillations are much smaller than those of the osculating elements. These actions can be considered as quasi-constants of motion for practical astronomical purposes. The first step for the computation of these actions is to construct the secular normal form, as already explained in Section 2.5.1, which allows one to identify in the mean semimajor axis the first approximate constant of motion. The second step is to construct a Birkhoff normal form of first order for the secular Hamiltonian (8.2). This step is detailed in the present section.

From the previous sections, we know that there are two possible integrable approximations of the secular Hamiltonian. The construction of the Birkhoff normal form depends on the chosen integrable approximation, but the formal scheme is the same in both cases. Therefore, in the following we will generically denote by $\mathcal{H}_0$ the integrable approximation, which can be either (8.13) or (8.25); moreover by $I_1, I_2, \varphi_1, \varphi_2$ we will denote the action–angle variables introduced for the integrable approximation, namely P', Q', p', q' for the linear approximation and J_1, J_2, ψ_1, ψ_2 for the Kozai approximation. Finally, by $\mathcal{H}_1$

we will generically denote the term of lowest order in the perturbation, i.e. $\mathcal{H}_{(4)}$ for the approach based on the linear approximation, or $\mathcal{K}_{(1)}$ for the one based on Kozai's approximation. In the linear approximation, the frequencies of the small body's longitudes of perihelion and node depend only on its semimajor axis; to have a more realistic approximation, we include in $\mathcal{H}_0$ the terms of $\mathcal{H}_1$ that are independent of the angles. In this way, $\mathcal{H}_0$ becomes nonlinear, while the average of $\mathcal{H}_1$ over all angles becomes equal to zero. In the Kozai approximation, $\mathcal{H}_0$ is already nonlinear and $\mathcal{H}_1$ has zero average over the angles, so that we do not need to perform any modification. In conclusion, we deal with a Hamiltonian of type

$$\mathcal{H} = \mathcal{H}_0(I_1, I_2, \Lambda_{g_k}, \Lambda_{s_k}) + \mathcal{H}_1(I_1, I_2, \varphi_1, \varphi_2, \varpi_k^*, \Omega_k^*) \tag{8.26}$$

where $\mathcal{H}_1$ is *small* (say of order η) with respect to $\mathcal{H}_0$ and has zero average over the angles, and $\mathcal{H}_0$ has the form

$$\mathcal{H}_0(I_1, I_2, \Lambda_{g_k}, \Lambda_{s_k}) = \sum_k (g_k \Lambda_{g_k} + s_k \Lambda_{s_k}) + \mathcal{W}_0(I_1, I_2) \ . \tag{8.27}$$

The function $\mathcal{W}_0$ is nonlinear in I_1, I_2, so that the frequencies $\omega_1 = \partial \mathcal{W}_0 / \partial I_1$, $\omega_2 = \partial \mathcal{W}_0 / \partial I_2$ are functions of I_1, I_2. Remember that the Hamiltonian (8.26) depends parametrically on the value of Λ (the quasi-constant action defined when computing the secular normal form; see Section 2.5), that is on the small body's *mean* semimajor axis.

To compute the Birkhoff normal form, $\mathcal{H}_1$ is first expanded in a Fourier series of the angles, as:

$$\mathcal{H}_1 = \sum_{j,l,\mathbf{m},\mathbf{n}} c_{j,l,\mathbf{m},\mathbf{n}}(I_1, I_2) \exp\left[\iota(j\varphi_1 + l\varphi_2 + \mathbf{m} \cdot \boldsymbol{\varpi}^* + \mathbf{n} \cdot \boldsymbol{\Omega}^*)\right] \ , \tag{8.28}$$

where $\boldsymbol{\varpi}^*$ and $\boldsymbol{\Omega}^*$ denote vectors with components ϖ_k^* and Ω_k^* respectively, the index k ranging over the considered number of planetary frequencies. In the approach based on the linear approximation, (8.28) contains only the terms of degree 4 in eccentricities and inclinations, so that, for the D'Alembert rules, it has only a finite number of harmonics; in the approach based on the Kozai approximation, the Fourier series on φ_1 (i.e. ψ_1) is infinite, so that one has to introduce some truncation on $|j|$ in order to retain in $\mathcal{H}_1$ only a finite number of terms. A seminumerical technique (see Lemaître and Morbidelli, 1994) allows one to work without explicit Fourier expansion in the angles, so that the truncation on $|j|$ is not needed. We will use here the explicit Fourier expansion for simplicity of explanation.

As shown in Chapter 2, the first-order Birkhoff normal form is obtained using the Lie series: new variables $I_1', I_2', \varphi_1', \varphi_2'$ are introduced with a transformation of type (2.6), with generating Hamiltonian χ given as a solution of equation (2.10). Using the expansion (8.28), χ can be written as:

$$\chi = \sum_{j,l,\mathbf{m},\mathbf{n}} -\iota \frac{c_{j,l,\mathbf{m},\mathbf{n}}(I_1', I_2')}{\mathbf{m}\cdot\mathbf{g} + \mathbf{n}\cdot\mathbf{s} + j\omega_1(I_1', I_2') + l\omega_2(I_1', I_2')} \exp\left[\iota(j\varphi_1' + l\varphi_2' + \mathbf{m}\cdot\boldsymbol{\varpi}^* + \mathbf{n}\cdot\boldsymbol{\Omega}^*)\right] , \tag{8.29}$$

where $\mathbf{g}$ and $\mathbf{s}$ are the vectors of planetary frequencies g_k and s_k respectively. The construction of the normal form fails if one of the denominators in (8.29) is close to zero; when this occurs, we say that there is a *secular resonance* among the planetary frequencies and the small body's frequencies. Because the approch based on the linear approximation and the one based on the Kozai approximation define perturbations $\mathcal{H}_1$ that have different Fourier expansions, the sets of secular resonances that prevent the construction of the first-order normal form are different in the two approaches. Similarly to what was discussed for the secular normal form (see Fig. 2.1), to guarantee that the denominators in (8.29) are not smaller than $\sqrt{\eta}$, the values of I_1 and I_2 of the small body must differ by at least $O(\sqrt{\eta})$ from the solutions of the equations

$$\mathbf{m}\cdot\mathbf{g} + \mathbf{n}\cdot\mathbf{s} + j\omega_1(I_1, I_2) + l\omega_2(I_1, I_2) = 0 \ . \tag{8.30}$$

Once χ is determined, the values of $I_1', I_2', \varphi_1', \varphi_2'$ that correspond to the values of $I_1, I_2, \varphi_1, \varphi_2$ of the asteroid can be computed to first order as:

$$\begin{aligned}
I_1 =& I_1' + \{I_1', \chi\} = I_1' - \tfrac{\partial\chi}{\partial\varphi_1'}(I_1', I_2', \varphi_1', \varphi_2', \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*) \\
I_2 =& I_2' + \{I_2', \chi\} = I_2' - \tfrac{\partial\chi}{\partial\varphi_2'}(I_1', I_2', \varphi_1', \varphi_2', \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*) \\
\varphi_1 =& \varphi_1' + \{\varphi_1', \chi\} = \varphi_1' + \tfrac{\partial\chi}{\partial I_1'}(I_1', I_2', \varphi_1', \varphi_2', \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*) \\
\varphi_2 =& \varphi_2' + \{\varphi_2', \chi\} = \varphi_2' + \tfrac{\partial\chi}{\partial I_2'}(I_1', I_2', \varphi_1', \varphi_2', \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*) \ .
\end{aligned} \tag{8.31}$$

However, these equations are in implicit form, because the generating function χ depends on the new variables. The solution of (8.31) is therefore computed by iterations, starting from $I_1' = I_1, I_2' = I_2, \varphi_1' = \varphi_1, \varphi_2' = \varphi_2$. If the derivatives of χ are small, the iterative method converges, and the solution values of $I_1', I_2', \varphi_1', \varphi_2'$ are called the *proper actions and angles* of the asteroid. The frequencies $\omega_1(I_1', I_2'), \omega_2(I_1', I_2')$ are the *proper frequencies.* Conversely, in the proximity of a secular resonance, where one of the denominators in (8.29) is small (typically smaller than $\sqrt{\eta}$), the derivatives of χ are large and the iterative method generally fails to converge.

The actions I_1' and I_2' (when computable) are precisely the new approximate constants of motion of the asteroid that we were looking for, at the beginning of this section. Instead of looking at the distribution of the asteroid

in the space of the quasi-invariant actions Λ, I'_1, I'_2, one usually prefers to first introduce more directly understandable quantities, called the *proper elements*, i.e. values of the semimajor axis, eccentricity and inclination related to the quasi-invariant actions. This is done as follows. The proper semimajor axis is identified with the mean semimajor axis, namely with $a_p = \Lambda^2/(\mathcal{G}M_\odot)$, where $\mathcal{G}$ is the gravitational constant and $M_\odot$ is the mass of the Sun. Then, in the approach based on the linear approximation, one assumes that the linear proper actions P', Q' are equal to the values of the proper actions I'_1, I'_2 (which is true on average) and computes the proper eccenticity and inclination using (8.14). In the approach based on the Kozai approximation, one assumes that the actions J_1 and H of the Kozai Hamiltonian are equal to the values of the proper actions I'_1, I'_2, thus identifying a specific cycle of eccentricity and inclination as a function of the argument of perihelion g; then one chooses as proper eccentricity and inclination some values of e and i that, together with a_p, unequivocally identify that cycle: for instance the values of e and i on the cycle when $g = 0$ or $g = 90°$, or alternatively the average of e and i over the cycle.

Hirayama (1918) was the pioneer of the linear approximation approach for the computation of proper elements: he used a secular normal form to first order in the mass of Jupiter, of which he considered only the linear part (thus skipping the computation of the Birkhoff normal form). Later Yuasa (1973) developed a more complete theory of asteroid secular motion, based on a secular normal form to second order in the mass of Jupiter, and retaining all the terms up to degree 4 in the eccentricities and inclinations. However Yuasa never used his theory to explicitely compute the asteroids' proper elements. This was done by Milani and Knežević (1990, 1992, 1994), who made Yuasa's theory more sophisticated and introduced the method based on the first-order Birkhoff normal form, explained in this section. In their most advanced algorithm Milani and Knežević (1994) took into account the perturbations given by Jupiter and Saturn, and computed the secular normal form to order 2 in the planetary masses, retaining all terms up to degree 4 in the eccentricities and inclinations; moreover they included terms accounting for the corrections of the asteroid's proper frequencies due to the terrestrial planets; for Jupiter's and Saturn's orbital elements, they assumed the linear Lagrange–Laplace theory, including all the terms related to the four giant planets' secular system, completed by the major terms coming from a nonlinear secular theory of planetary motion.

The approach based on the Kozai Hamiltonian was first introduced by Williams (1969). Williams did not work in the framework of Hamiltonian formalism, but his algorithm for the computation of proper elements is basi-

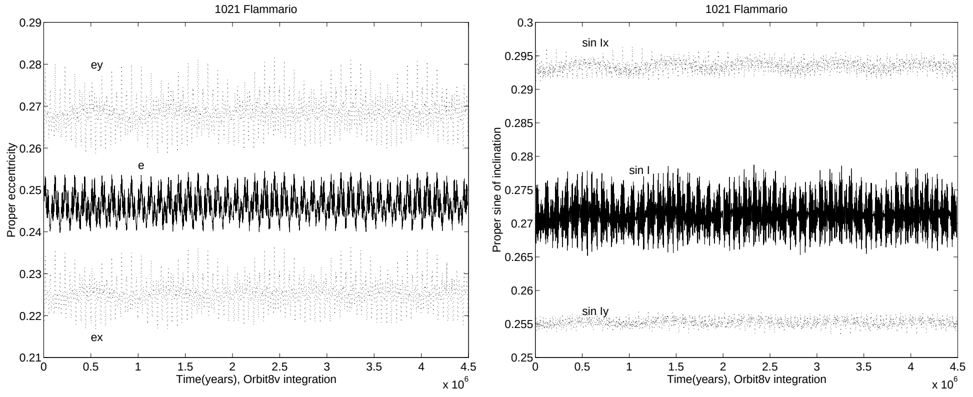


Figure 8.5: Time evolution of the proper elements of the asteroid 1021 Flammario. The labels e and I denote the proper eccentricity and proper inclination computed using Milani and Knežević's algorithms; ex, Ix are the proper eccentricity and inclination computed using Lemaître and Morbidelli's algorithm, defined as the values of e and i when $g = 0$ on the Kozai cycle identified by the value of the proper actions; ey, Iy are the same as ex, Ix, but defined as the values of e and i when $g = 90°$ on the Kozai cycle. For this asteroid (inclination $\sim 15.5°$) the accuracy of the different sets of proper elements – defined as the root mean square of their dispersion with respect to their average value – is comparable. Reprinted from Fig. 2 of Knežević *et al.* (1995), with permission from Astronomy and Astrophysics.

cally equivalent to that explained above; in particular his secular equations are equivalent to those arising from a secular normal form computed up to first order in the mass of Jupiter. Lemaître and Morbidelli (1994) revisited Williams' algorithm in the framework of Hamiltonian theory, thus introducing the procedure based on the use of Arnold's action–angle variables for the Kozai Hamiltonian, explained in this section. Moreover, they started from a secular normal form computed up to second order in the planetary masses, and took into account also the perturbations of Saturn on the small body. For Jupiter's and Saturn's orbital elements they assumed Lagrange–Laplace linear theory with forcing frequencies $g_5, g_6, g_7, s_6, s_7, s_8$, plus the leading nonlinear term with frequency $2g_6 - g_5$.

The accuracy of proper elements has been tested using numerical integrations: a number of asteroids have been integrated over several million years, and, at each output time, their proper elements have been computed. In principle, the proper elements should be constants of motion, so that their values

should be the same at every output time; in reality, the values of the proper elements oscillate with time (see Fig. 8.5), due to the terms that have not been taken into account in the low-order normal forms used for their computation. Therefore, the accuracy of the proper elements is measuread as the root mean square (rms) of their dispersion with respect to their average value over the entire integration timespan. For the asteroids with low eccentricity and inclination which are not close to secular resonances the accuracy of Milani and Knežević's proper elements is excellent: the rms of the proper eccentricity and of the sine of the proper inclination are typically much smaller than 10^{-3} (Milani and Knežević, 1994). However, the accuracy degrades for asteroids with larger eccentricity and/or inclination; this is because their Hamiltonian is a truncated power series expansion in e and i, and because the linear approximation assumes that e and i are constant during the precession of the argument of perihelion, which is a coarse approximation for bodies not too far from the Kozai resonance (see Fig. 8.1). For bodies with $i \sim 15$ degrees, the rms of Milani and Knežević's proper elements become of order 5×10^{-3} for the eccentricity and 10^{-3} for the inclination, i.e. comparable to the typical accuracy of the proper elements by Lemaître and Morbidelli (see Knežević *et al.*, 1995, and Fig. 8.5). The accuracy of the latter is basically the same for all asteroids, whatever their eccentricity and inclination, because no expansion is done in powers of e and i and because the Kozai dynamics is accurately taken into account. In principle, Lemaître and Morbidelli's proper elements can be computed also for asteroids librating inside the Kozai resonance, but this fact has little importance for practical applications because only a very limited number of these asteroids are known. However, the proper elements computed by Lemaître and Morbidelli never achieve the best accuracy of those by Milani and Knežević, because the quadratic and cubic terms in the planetary eccentricities and inclinations (i.e. $\mathcal{K}_{(2)}, \mathcal{K}_{(3)}$) are not taken into account.

8.3.1 Asteroid families

Figure 8.6 finally shows the distribution of the asteroids with respect to the proper semimajor axis, inclination and eccentricity computed by Milani and Knežević. Several regions of overdensity are immediately evident. A rigorous analysis of the asteroid distribution with respect to the proper a, e and i has been done with two independent classification methods by Zappalà *et al.* (1990) and Bendjoya *et al.* (1991), leading to similar results (Zappalà *et al.*, 1995): there are 32 regions of statistically significant overdensity in the distribution of the asteroids with proper inclination smaller than 17.5 degrees. There is a general consensus that these regions of overdensity are the result of the break up of a parent body due to collision with another (generally

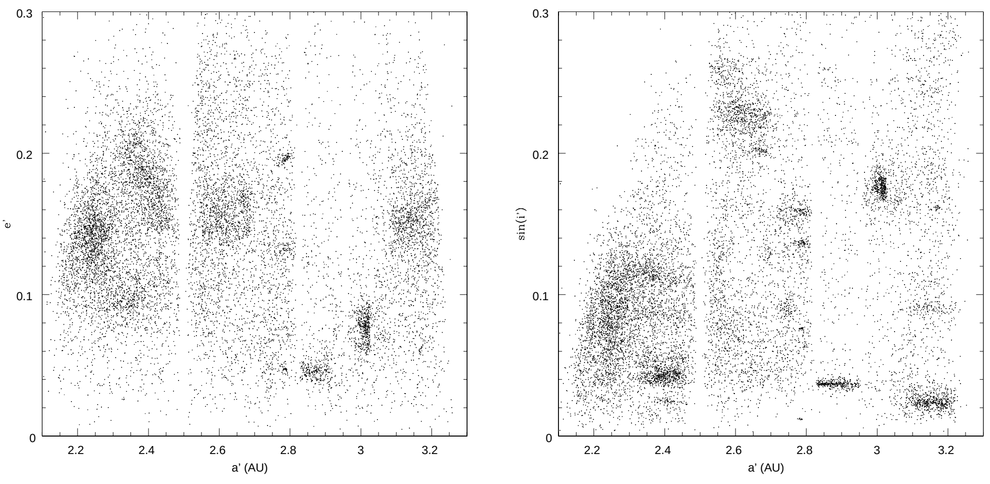


Figure 8.6: The distribution of 12,487 asteroids in the proper element space. a', e' and i' denote here the values of the proper semimajor axis, eccentricity and inclination computed by Milani and Knežević. Reprinted from Fig. 1 of Zappalà *et al.* (1995), with permission from Academic Press.

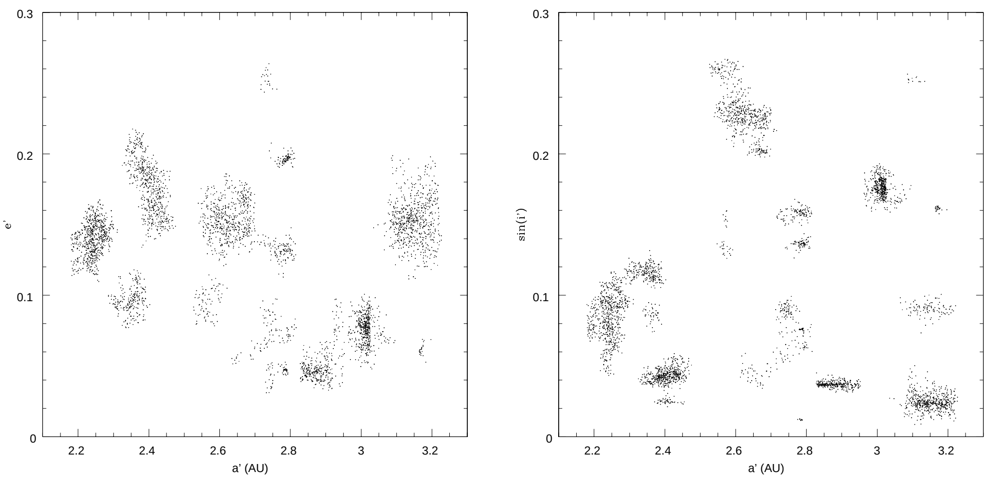


Figure 8.7: The same as Fig. 8.6, but only the members of the 32 identified families are plotted. Reprinted from Fig. 12 of Zappalà *et al.* (1995), with permission from Academic Press.

smaller) asteroid: for this reason they are generally called *asteroid families*. The asteroid families identified with statistical reliability with proper inclination smaller than 17 degrees are shown in Fig. 8.7. At larger inclination, six more families have been identified using the proper elements by Lemaître and Morbidelli by Forzoni Accolti (1995): three are in the Hungaria region (the region populated by the asteroids with $a \sim 1.9$ AU, $i \sim 22°$), two are in the Phocea region ($a \sim 2.35$ AU, $i \sim 24°$), and the last one is associated with the large asteroid 2 Pallas ($a = 2.77$ AU, $e = 0.23$, $i = 34.8°$ in Fig. 2).

The high accuracy of Milani and Knežević's proper elements allows astronomers to study in detail the distribution of the asteroids inside the family, thus reconstructing the velocity field that characterized the ejection of the fragments from their parent body (Cellino *et al.*, 1999). This in turn allows one to deduce several properties of the collisional physics of asteroids, namely for a range of sizes and impact velocities that are orders of magnitude larger than those characterizing laboratory experiments.

8.4 Secular resonances

The locations of secular resonances are simply computed as the sets of the actions Λ, I_1, I_2 such that the frequencies of the integrable Hamiltonian (8.27) satisfy relationships of the type (8.30), with coefficients $\mathbf{m}, \mathbf{n}, j$ and l such that the corresponding harmonics $\exp \iota(j\varphi_1 + l\varphi_2 + \mathbf{m} \cdot \boldsymbol{\varpi}^* + \mathbf{n} \cdot \boldsymbol{\Omega}^*)$ satisfy the D'Alembert rules and effectively appear in the Fourier expansion of the perturbation. In the approach based on the linear approximation, the use of (8.27) to locate linear secular resonances is done with some abuse. In fact, in the presence of linear resonances, one of the amplitudes (8.8) of the transformation (8.7) is infinite, so that (8.27) cannot be constructed following the perturbation approach described in Section 8.1. In this cases, the integrable "approximation" (8.27) is defined by simply neglecting all harmonic terms in $\mathcal{H}_{(2)}$ and $\mathcal{H}_{(4)}$.

Once the values of the actions Λ, I_1, I_2 corresponding to a secular resonance are determined, the related values of semimajor axis, eccentricity and inclination are computed as for the proper elements case.

In the asteroid belt, the location of secular resonances has been computed by Milani and Knežević (1990, 1992, 1994), following the approach based on the linear approximation, and by Williams and Faulkner (1981) and subsequently by Morbidelli and Henrard (1991a), following the approach based on the Kozai approximation. Figure 8.8 shows Milani and Knežević's result: the location of secular resonances is given with respect to proper semimajor axis and inclination, for a value of the proper eccentricity fixed equal to 0.1; only

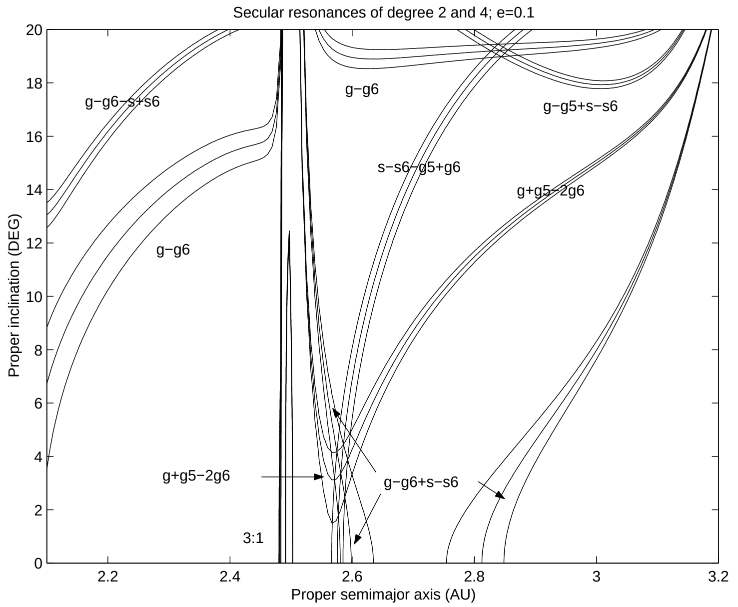


Figure 8.8: The location of secular resonances and their estimated width for proper eccentricity equal to 0.1. Only the secular resonances whose corresponding harmonics appear in $\mathcal{H}_{(2)}$ or $\mathcal{H}_{(4)}$ are shown. Reprinted from Fig. 5 of Milani and Knežević (1990), with permission from Kluwer Academic Publishers.

the linear resonance $g - g_6$ and the resonances whose corresponding harmonics appear in the main perturbation term $\mathcal{H}_{(4)}$ are represented. The location of the resonances for other values of the proper eccentricity can be found in the above-quoted Milani and Knežević papers. In the authors' notation, g and s are respectively the frequencies of the longitudes of perihelion and node which should be identified with $-\omega_1$ and $-\omega_2$ in the notation of formula (8.30).[1] For each resonance, the central curve shows the exact location, where the corresponding relation (8.30) is precisely satisfied, while the two curves on each side show the locations where the same frequency combination is equal to $\pm 1''/y$, in the case of the $g - g_6$ resonance, or $\pm 0.5''/y$, in the case of the other resonances. These arbitrary values have been chosen to give a qualitative indication of the widths of the resonances. Note that the resonances appear to accumulate and assume a vertical location close to the 3:1 mean motion resonance, approxi-

[1]The reader should not confuse the argument of perihelion and the frequency of the longitude of perihelion, which are both denoted by g in this chapter, but always in a different context.

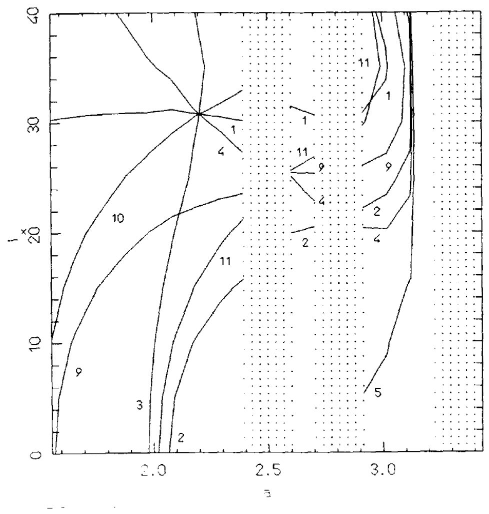


Figure 8.9: The location of secular resonances for proper eccentricity equal to 0.1 and outside of the Kozai resonance. Only the secular resonances whose corresponding harmonics appear in $\mathcal{K}_{(1)}$ or $\mathcal{K}_{(2)}$ are reported. The resonance enumeration is the following: 1: $g = g_5$ (ν_5); 2: $g = g_6$ (ν_6); 3: $s = s_6$ (ν_{16}); 4: $g + s = g_5 + s_6$; 5: $g + s = g_6 + s_6$; 9: $2g = g_5 + g_6$; 10 $g - s = g_5 - s_6$; 11: $g - s = g_6 - s_6$. Here, g and s are the proper frequencies of the longitude of perihelion and of the longitude of node of the asteroid, to be identified respectively with $\omega_1 + \omega_2$ and ω_2 in the notation of (8.30). The dotted bands denote the regions close to the main mean motion resonances where the secular normal form is singular. Reprinted from Fig. 9 of Morbidelli and Henrard (1991a), with permission from Kluwer Academic Publishers.

mately at 2.5 AU. This is an artifact of the computation method: the secular normal form, in fact, is singular at the location of mean motion resonances because of the presence of the small denominators introduced in (2.41) for the elimination of the corresponding harmonics. The appropriate way to compute the location of secular resonances inside/close to mean motion resonances will be explained in Chapter 11.

Figure 8.9 shows the location of secular resonances according to Morbidelli and Henrard (1991a), which is computed over a more extended range of inclinations, thanks to the fact that in the approach based on the Kozai approximation the computation of the asteroid's secular frequencies does not lose accuracy at large i. All resonances whose corresponding harmonics appear in the perturbation terms $\mathcal{K}_{(1)}$ and $\mathcal{K}_{(2)}$ are shown. As in Fig. 8.8, the computa-

tion is done for a proper eccentricity equal to 0.1; here, proper eccentricity and inclination are defined as the values of e and i assumed when the argument of perihelion is equal to zero, on the cycle defined by the Kozai Hamiltonian. The location of secular resonances is computed with reference to the secular normal form of order 2 in Jupiter's mass; the results are therefore improved with respect to those by Williams and Faulkner (1981), who worked in an equivalent way, but starting from a secular normal form that was computed only to first order in Jupiter's mass. The dotted bands around 2.5, 2.8 and 3.3 AU denote "forbidden regions" where the secular normal form is singular due to the presence of the 3:1, 5:2 and 2:1 mean motion resonances, respectively. Figure 8.9 concerns only orbits that are outside of the Kozai resonance, for which the argument of perihelion circulates; however, the approach based on the Kozai approximation also allows the computation of the location of secular resonances in the regions where the argument of perihelion librates, inside the Kozai resonance. There, the asteroid's longitude of perihelion precesses in the opposite direction with respect to the planets' longitudes of perihelia, so that low-order secular resonances involving the perihelia cannot exist; as a consequence, only the $s - s_6$ resonance has been found inside the Kozai resonance by Morbidelli and Henrard (1991a).

Michel and Froeschlé (1997) extended Morbidelli and Henrard's computation to the low-eccentricity orbits in the inner Solar System (0.5–2 AU region) which do not intersect the orbits of the terrestrial planets. Moreover they devoted their attention also to the resonances with the terrestrial planets' secular frequencies, usually neglected in asteroid belt computations. The inner Solar System turns out to be crowded with these resonances, which play a nonnegligible role in the dynamical evolution of near-Earth asteroids (NEAs) (Michel, 1997).

Comparing Fig. 8.9 with the left panel of Fig. 2, it is evident that the distribution of the asteroids is sculpted by the main secular resonances. The $g - g_6$ resonance (also called the ν_6 resonance) bounds the distribution of the principal population of the asteroids in the belt, while the groups of Hungaria and Phocea (respectively $a \sim 1.9$ AU., $i \sim 22°$ and $a \sim 2.35$ AU, $i \sim 24°$) seem to be confined also by the $g - g_5$ (ν_5) and $s - s_6$ (ν_{16}) resonances. As will be detailed below, this is a consequence of the large eccentricity or inclination variations that these resonances may force, eventually destabilizing the asteroid's motion; the ν_5, ν_6 and ν_{16} resonances have therefore created large gaps in the asteroids' distribution. Conversely, comparison between Figs. 8.8 and 2b reveals that the main belt is crossed by several secular resonances of higher order, but which do not seem to affect the asteroids' distribution. Actually, Fig. 8.8 reminds us of the Nekhoroshev structure sketched in Fig. 6.7a, where

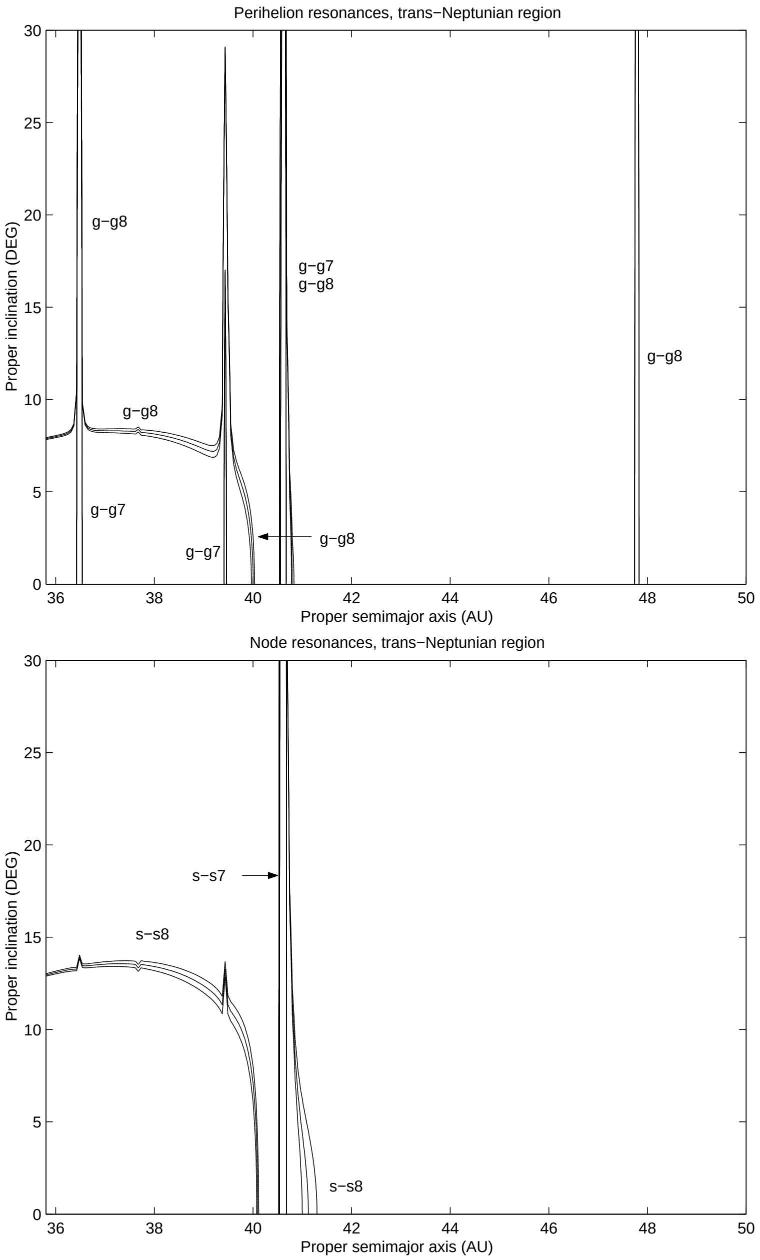


Figure 8.10: The location of the linear secular resonances in the Kuiper belt for proper eccentricity equal to 0.1. Adapted from Fig. 5 of Knežević *et al.* (1991), with permission from Academic Press.

resonances cross each other at resonant nodes, but do not completely overlap. This structure suggests that asteroids in nonlinear secular resonances, despite of their possible chaotic motion, should be "stable" for very long times. In fact, Milani and Knežević (1990, 1992, 1994) have numerically integrated over several million years the evolution of a number of asteroids in these secular resonances, without observing macroscopic drift of their proper elements. The example case is that of the asteroid 221 Eos, in the $g + s - g_6 - s_6$ resonance: this asteroid is the parent body of one of the most populated families, of which several members share its same chaotic behavior due the presence of the resonance; however the family is still very compact and easily identifiable, because of the absence of macroscopic diffusion. The concept of "practical" stability, introduced by Nekhoroshev theory, finds in the Eos family an enlightening practical example.

In the Kuiper belt, the location of secular resonances has been computed by Knežević *et al.* (1991), following the same approach used by Milani and Knežević for the asteroid belt (but including also the perturbations of Uranus and Neptune on the small bodies). In this case, the results are expected to be valid over a larger range of inclinations than in the asteroid belt, because the Kozai resonance never appears at low eccentricity, as seen in Fig. 8.2. Figure 8.10 shows the resulting location of the linear secular resonances. The accumulation of the perihelion resonances around 36.5 AU, 39.5 AU and 47.5 AU is an artifact of the singularities of the secular normal form that are due to the presence of the 3/4, 2/3 and 1/2 mean motion resonances with Neptune. Apart from these singular situations, three regions in the Kuiper belt are mostly affected by secular resonances: The 40–42 AU region is affected by the $g - g_8$ (alias ν_8), $s - s_7$ (ν_{17}) and $s - s_8$ (ν_{18}) resonances; the region at about 36 AU is crossed by the $g - g_7$ (ν_7) and ν_{17} resonances; in addition, in the region between 36.5 and 39.5 AU both the ν_8 and ν_{18} resonances are present for inclinations between 10 and 15 degrees. The outer part of the Kuiper belt, beyond 42 AU, does not present any linear secular resonance: there, the precession frequencies of the orbits of the Kuiper belt objects are slower than those of all the outer planets, so that only higher-order resonances could be found.

The location of secular resonances in the regions among the giant planets has also been determined by Knežević *et al.* (1991); the corresponding figures can be found in that paper.

8.4.1 Secular resonant dynamics

The dynamics of bodies in secular resonances can be analytically studied by constructing the resonant normal form, following the approach explained in

Section 2.3.1. For this purpose, it is more suitable to follow the approach based on the Kozai approximation. In this section we will concentrate on the case of *secular resonances of first order*, which are those whose corresponding harmonics appear in the first-order perturbation term $\mathcal{K}_{(1)}$.

We start from the Hamiltonian (8.26) which, in the case of the Kozai approach, can be written as:

$$\mathcal{H} = \sum_k (g_k \Lambda_{g_k} + s_k \Lambda_{s_k}) + \mathcal{K}_{(0)}(J_1, J_2) + \mathcal{K}_{(1)}(J_1, J_2, \psi_1, \psi_2, \varpi_k^*, \Omega_k^*) \ , \quad (8.32)$$

where J_1, ψ_1, J_2, ψ_2 are the action–angle variables for the Kozai Hamiltonian, introduced as explained in Section 8.2.3. Remember that the angle ψ_2 is "close" to the longitude of node h of the small body; the angle ψ_1 is "close" to the argument of perihelion g in the case where the latter circulates; conversely, ψ_1 is related to the angular position relative to the libration center in the case where g librates in the Kozai resonance. The frequencies of ψ_1 and ψ_2 are respectively $\omega_1 = \partial \mathcal{K}_{(0)}/\partial J_1$ and $\omega_2 = \partial \mathcal{K}_{(0)}/\partial J_2$. Because of the D'Alembert rules, the perturbation $\mathcal{K}_{(1)}$ must have Fourier expansion:

$$\mathcal{K}_{(1)} = \sum_{k,l} \left[c_{l,k}(J_1, J_2) \cos(l\psi_1 + \psi_2 - \varpi_k^*) + d_{l,k}(J_1, J_2) \cos(l\psi_1 + \psi_2 - \Omega_k^*) \right] \ . \quad (8.33)$$

In the case where the argument of perihelion circulates, in (8.33) the coefficients $c_{l,k}$ with even l and $d_{l,k}$ with odd l are equal to zero. In these variables, the harmonic related to the $g - g_k$ resonance is $\cos(\psi_1 + \psi_2 - \varpi_k^*)$, while the harmonic related to the $s - s_k$ resonance is $\cos(\psi_2 - \Omega_k^*)$.

To fix notation, let us consider a body that is close to a single resonance whose related harmonic is $\cos(l_*\psi_1 + \psi_2 - \varpi_{k_*}^*)$. The case where the body is close to a single resonance with harmonic $\cos(l_*\psi_1 + \psi_2 - \Omega_{k_*}^*)$ is trivially analogous. As shown in Section 2.3.1, the first-order resonant normal form is constructed using the Lie series: new variables $J_1', \psi_1', J_2', \psi_2', \Lambda_{g_k}', \Lambda_{s_k}'$ (called *semiproper*) are introduced by a transformation of type (2.6), with generating Hamiltonian χ given as the solution of the equation

$$\mathcal{K}_{(1)} + \{\mathcal{K}_{(0)}, \chi\} = c_{l_*,k_*}(J_1', J_2') \cos(l_*\psi_1' + \psi_2' - \varpi_{k_*}^*) \ . \quad (8.34)$$

This gives

$$\chi = \sum_{(k,l)\neq(k_*,l_*)} \frac{c_{l,k}(J_1', J_2')}{l\omega_1(J_1', J_2') + \omega_2(J_1', J_2') - g_k} \sin[l\psi_1' + \psi_2' - \varpi_k^*]$$
$$+ \sum_{k,l} \frac{d_{l,k}(J_1', J_2')}{l\omega_1(J_1', J_2') + \omega_2(J_1', J_2') - s_k} \sin[l\psi_1' + \psi_2' - \Omega_k^*] \ , \quad (8.35)$$

where none of the denominators is close to zero, by the assumption that the body is not close to any resonance, other than the considered one. Note that, because the generating Hamiltonian χ does not depend on $\Lambda'_{g_k}, \Lambda'_{s_k}$, the planetary angles ϖ_k^* and Ω_k^* are not changed by the transformation.

To first order (i.e. neglecting the higher-order terms introduced by the Lie series), the Hamiltonian in the new variables is reduced to

$$\mathcal{H} = \sum_k (g_k \Lambda'_{g_k} + s_k \Lambda'_{s_k}) + \mathcal{K}_{(0)}(J'_1, J'_2) + c_{l_*,k_*}(J'_1, J'_2) \cos(l_* \psi'_1 + \psi'_2 - \varpi^*_{k_*}) \ . \tag{8.36}$$

Because this Hamiltonian is independent of ϖ_k^* ($k \neq k_*$) and Ω_k^*, the conjugate actions are constants of motion. Then, dropping the constant terms $\sum_{k\neq k_*} \Lambda'_{g_k}$ and $\sum_k s_k \Lambda'_{s_k}$, and introducing the new canonical variables

$$\begin{aligned} S \ &= J'_2 \ , & \sigma \ &= l_* \psi'_1 + \psi'_2 - \varpi^*_{k_*} \\ C \ &= J'_1 - l_* J'_2 \ , & &\psi'_1 \\ \tilde{\Lambda}_{g_{k_*}} &= \Lambda'_{g_{k_*}} + J'_2 \ , & &\varpi^*_{k_*} \ , \end{aligned} \tag{8.37}$$

the Hamiltonian becomes

$$\mathcal{H} = \mathcal{K}_{(0)}(S, C) - g_{k_*} S + c_{l_*,k_*}(S, C) \cos \sigma \ , \tag{8.38}$$

where also the constant term $g_{k_*} \tilde{\Lambda}_{g_{k_*}}$ has been omitted. The Hamiltonian (8.38) is trivially integrable, since it depends only on one angle. Note that (8.38) has the same form as (4.2), i.e. the standard form for an integrable single resonance normal form. The angle σ is the critical angle of the considered secular resonance.

At this point, the recipe to compute the resonant evolution of a body is conceptually simple. From the initial action–angle variables $\mathbf{J} \equiv (J_1, J_2), \boldsymbol{\psi} \equiv (\psi_1, \psi_2)$ of the body, one first computes the values of the *semiproper action–angle variables* $\mathbf{J}' \equiv (J'_1, J'_2), \boldsymbol{\psi}' \equiv (\psi'_1, \psi'_2)$ by iteratively solving the implicit equations $\mathbf{J} = \mathbf{J}' + \{\mathbf{J}', \chi(\mathbf{J}', \boldsymbol{\psi}')\}$, $\boldsymbol{\psi} = \boldsymbol{\psi}' + \{\boldsymbol{\psi}', \chi(\mathbf{J}', \boldsymbol{\psi}')\}$, with χ given by (8.35), similarly to what is done in (8.31) for computing the proper action–angle variables of nonresonant asteroids. Then, using (8.37) one computes the values S_b, σ_b, C_b of S, σ and C for the body. These are used as initial conditions for computing the evolution according to the integrable Hamiltonian (8.38). To obtain a global view of the resonant dynamics in the S, σ plane on which the body evolves, one can simply plot on the surface $C = C_b$ the level curves of (8.38); in this diagram, the evolution of the body is represented by the level curve passing through the point S_b, σ_b. In practical applications, it is

often preferable to represent the dynamics in variables that are more directly readable than the actions S and C. For this purpose, for every value of S and C one computes the semiproper actions J_1', J_2' using the inverse of (8.37) and then defines semiproper eccentricity and inclination in a way analogous to that used to relate the proper actions to the proper eccentricity and inclination. More precisely, one assumes that the actions J_1 and H of the Kozai Hamiltonian are equal to the values of the semiproper actions J_1', J_2', thus identifying a specific cycle of eccentricity and inclination as a function of the argument of perihelion g; then the semiproper eccentricity and inclination may be defined as the values of e and i on the cycle when $g = 0$. Of course, the evolution of the semiproper eccentricity and inclination represents just the skeleton of the real evolution of the body: the body's e and i oscillate with respect to the semiproper eccentricity and inclination, due to the effects of the Kozai dynamics and of all the nonresonant terms in $\mathcal{K}_{(1)}$ that have been averaged out in the construction of the resonant normal form.

Although simple in principle, the above sketched procedure is technically complicated. Thus, Nakai and Kinoshita (1985) and Yoshikawa (1987), who were the first to study the dynamics in the ν_{16} and ν_6 secular resonances, introduced a number of simplifications. First, instead of introducing Arnold action–angle variables for the Kozai Hamiltonian as in Section 8.2.3, they simply averaged $\mathcal{K}_{(0)}(G, H, g)$ over the argument of perihelion g. In the notation of this chapter, this is equivalent to reducing the transformation (8.23) to the identity $J_1 = G$, which is an accurate approximation only for small eccentricities and inclinations, where the cycles of the Kozai Hamiltonian in polar coordinates G, g are close to circles. Second, they computed the average of $\mathcal{K}_{(1)}$ over the nonresonant harmonics without computing the generating Hamiltonian χ. Therefore, they identified the semiproper actions J_1', J_2' with J_1, J_2, i.e. with the Delaunay actions G, H. This approximation was motivated by the fact that the authors were not interested in computing the secular evolution of specific asteroids, but rather aimed to explore the general dynamical properties of the resonances. Third, in the work of Yoshikawa (1987) the secular normal form was expanded in a power series of the eccentricity and inclination of the asteroid, and truncated to degree 4. We do not detail here the results obtained by Nakai and Kinoshita and by Yoshikawa, for which the reader can directly consult the original papers. Recall that their results are quantitatively accurate only for small to moderate eccentricities and inclinations. In some cases, also qualitative differences with the real dynamics can occur: for instance in Yoshikawa's theory the libration center of the ν_6 resonance at large inclination turns out to be rotated by 180 degrees with respect to the real one, as discussed by Morbidelli and Henrard (1991b).

The approximations introduced by Nakai and Kinoshita and by Yoshikawa, have the merit of showing a general property of secular resonances that otherwise does not appear evident in the action–angle formalism detailed in this section. By identifying J_1', J_2' with G, H, it turns out from (8.37) that for perihelion secular resonances of type $g - g_k = 0$ (which correspond to $l_* = 1$) one has $S = H$ and $C = G - H$, while for nodal resonances of type $s - s_k = 0$ (which correspond to $l_* = 0$ and Ω_k^* replacing ϖ_k^*) one has $S = H, C = G$. Recalling that $G - H \sim i^2$ and that both the semimajor axis and C are constant, this implies that perihelion resonances preserve the inclination and force a change in the eccentricity, while nodal resonances keep e constant and force the evolution of the inclination. Of course, this holds only when the identification $J_1' = G, J_2' = H$ is a good approximation, i.e. for small eccentricities and inclinations.

The approach detailed in this section was introduced and applied without simplification by Morbidelli (1993a). Figure 8.11 shows an example concerning the asteroid 945 Barcelona, in the ν_5 resonance (semimajor axis equal to 2.64 AU). The panel on the left shows the global picture of the dynamics on the surface $C = C_{\rm b}$, with $C_{\rm b}$ being the value of C for Barcelona. The dynamics is illustrated in the coordinates $x = e_{\rm sp}\cos\sigma_5, y = e_{\rm sp}\sin\sigma_5$, where $e_{\rm sp}$ is the semiproper eccentricity and σ_5 is the critical resonant angle, namely $\sigma_5 = \psi_1' + \psi_2' - \varpi_5^*$ in the notation of (8.37). The black dot corresponds to the present position of Barcelona. As one sees, the portrait of the resonant dynamics is very similar to that of a pendulum in polar coordinates, shown in Fig. 4.1b. In particular, one recognizes a stable equilibrium point at $\sigma_5 = 0$, surrounded by banana-shaped curves, which represent the librations in polar coordinates. According to the diagram and to the present position of Barcelona, the asteroid is expected to librate around $\sigma_5 = 0$, its semiproper eccentricity passing from approximately 0.15 to 0.2. It is very instructive to compare this theoretical result with a numerical integration, such as that done by Scholl and Froeschlé (1990) and shown in the two panels on the right side of Fig. 8.11. The top and bottom panels show respectively the forward and backward integrations, both for 1 My. The coordinates are $\Psi_1^J = [2(1-\sqrt{1-e^2})]^{1/2}\cos(\varpi - \varpi_{\rm J}) \sim e\cos(\varpi - \varpi_{\rm J})$, $\Psi_2^J = [2(1-\sqrt{1-e^2})]^{1/2}\sin(\varpi - \varpi_{\rm J}) \sim e\sin(\varpi - \varpi_{\rm J})$, where $e, \varpi, \varpi_{\rm J}$ are respectively the osculating eccentricity, the longitude of perihelion of the asteroid and the longitude of perihelion of Jupiter. The forward and backward numerical integrations show banana-shaped "bands", which are essentially made of an "epicyclic" motion superposed on an almost semicircular arc of evolution. The "epicycles" are associated with the precession of the argument of perihelion g, i.e. to the Kozai dynamics. The semiproper eccentricity $e_{\rm sp}$ used to

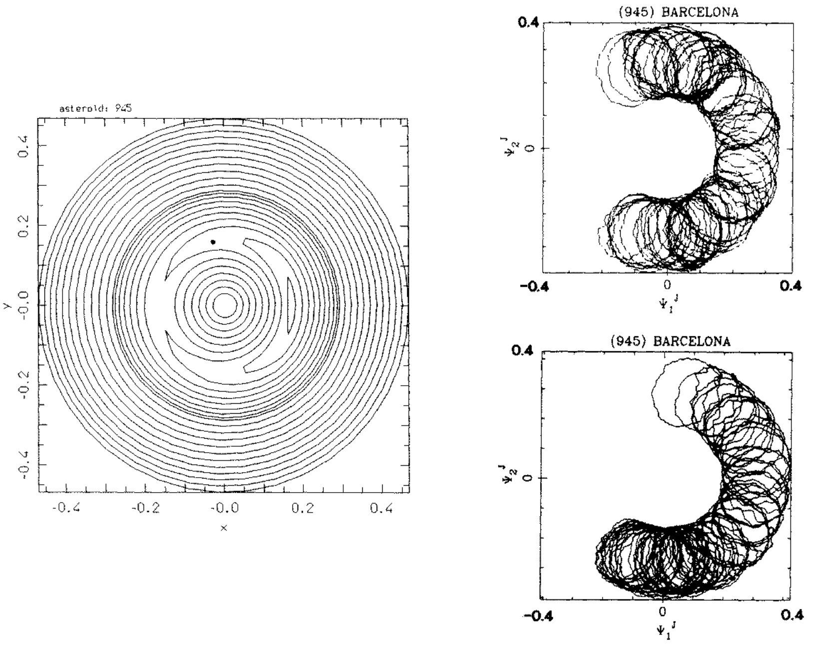


Figure 8.11: Left: the dynamical portrait of the ν_5 secular resonance, computed for the values of semiproper actions of the asteroid 945 Barcelona. The black dot denotes the present position of the asteroid; the coordinates are explained in the text. Right: the future (top) and past (bottom) evolution of the asteroid according to Scholl and Froeschle (1990). See text for discussions. Reprinted from Fig. 1 of Morbidelli (1993a), with permission from Academic Press.

represent the secular resonant motion in the left panel corresponds to the value of e when $g = 0$, i.e. to the minimal value of the eccentricity on each "epicycle". Therefore, one should compare the expected evolution of Barcelona in the $e_{\rm sp}, \sigma_5$ plane with the inner edge of the "bands" drawn by numerical simulation on the Ψ_1^J, Ψ_2^J plane. Doing so, one sees that the forward and backward integration each correspond roughly to half of a libration cycle in the ν_5 resonance, and that Barcelona cuts the $\Psi_2^J = 0$ axis approximately at $e_{\rm sp} = 0.15$ in the forward integration and $e_{\rm sp} = 0.2$ in the backward integration, in very good agreement with the result obtained through the computation of the secular resonant normal form. Note that in the numerical simulation the total variation of the osculating eccentricity is between 0.15 and 0.4, and that most of it is due to the Kozai dynamics, not to the ν_5 resonant dynamics. In re-

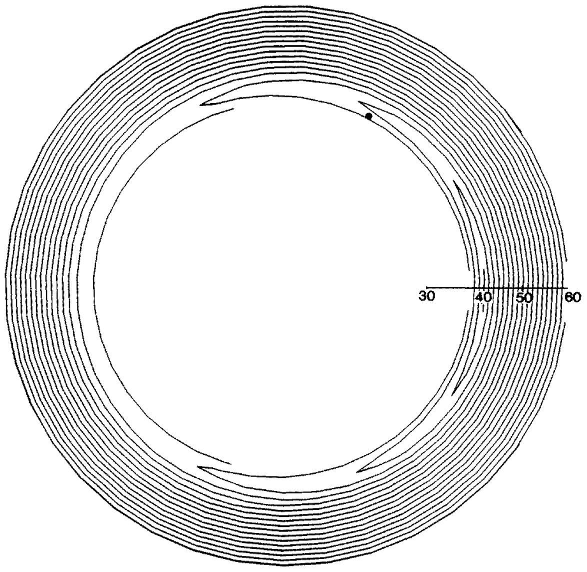


Figure 8.12: The dynamical portrait of the ν_{16} secular resonance, inside the Kozai resonance, computed for the values of semiproper actions of the asteroid 2335 James. The black dot denotes the present position of the asteroid. The coordinates are explained in the text. Reprinted from Fig. 5 of Morbidelli (1993a), with permission from Academic Press.

ality, the amplitude of variation of the eccentricity forced by the ν_5 secular resonance is in general quite limited (~ 0.05 in this case). This is due to the fact that, in the semimajor axis range corresponding to the asteroid belt, the coefficient of the harmonic $\cos(\psi_1' + \psi_2' - \varpi_5^*)$ is small, because it vanishes and changes sign for a value of the inclination ($\sim 30°$) that is close to that characterizing the location of the ν_5 resonance (see Morbidelli and Henrard, 1991b). The general scarcity of asteroids corresponding to the ν_5 resonance is therefore related more to the Kozai dynamics (responsible for the general depletion of the asteroid belt at large inclination by pumping the eccentricity to planet-crossing values), rather than to the ν_5 resonance itself.

The perturbation approach discussed in this section is valid also for studying the dynamics of secular resonances *inside* the Kozai resonance, i.e. for orbits with librating argument of perihelion. As an example, Fig. 8.12 is drawn for the asteroid 2335 James ($a \sim 2.12$ AU), which is in the ν_{16} resonance inside the Kozai resonance. The secular resonant dynamics is represented in polar coordinates: the radius is the semiproper inclination $i_{\rm sp}$, here defined as

the minimal value assumed by the inclination over the Kozai libration cycle (corresponding to one of the two intersections of the cycle with the $g = 90°$ axis); the angle is the critical one of the ν_{16} resonance, rotated by 180 degrees, namely $\sigma_{16} = \psi'_2 - \Omega^*_6 + \pi$ (the phase π has been added so that σ_{16} is close to the difference $\Omega - \Omega_\mathrm{J}$ between the osculating longitudes of node of the asteroid and of Jupiter; see formula 7.10 and Table 7.3). The black dot in Fig. 8.12 denotes the present position of the asteroid in these coordinates. The level curves of the secular resonant normal form (8.38) are plotted only in the region of libration of the argument of perihelion, and they are cut when they encounter the Kozai separatrix. As one sees from the plot, the asteroid James librates around $\sigma_{16} = 0$, its semiproper inclination passing from approximately 37 to 42 degrees. Its librational curve hits the Kozai separatrix at $i_\mathrm{sp} \sim 37°, \sigma_{16} = 0$, so that the asteroid may escape from the Kozai resonance, its argument of perihelion starting to circulate.

As shown in Fig. 8.9, the ν_{16} resonance is present in the asteroid belt also outside of the Kozai resonance; in this region its amplitude is much larger than inside the Kozai resonance and, as first shown by Nakai and Kinoshita (1985), it may exceed 15 degrees. Moreover, the stable libration center of the ν_{16} resonance is at $\sigma_{16} = 180°$, i.e. rotated by 180 degrees with respect to the Kozai resonant case of Fig. 8.12. The absence of asteroids between the Hungaria and the Phocea groups (see Fig. 2) can be explained by the presence of the ν_{16} resonance. In fact, asteroids in this location would suffer oscillations of their inclinations that are large enough to take them to interact either with the ν_6 resonance (at lower inclination) or with the ν_5 and Kozai resonances (at higher inclination), thus subsequently increasing their eccentricity up to at least Mars-crossing values.

In long-term numerical integrations (Froeschlé *et al.*, 1991), the asteroid 2335 James shows fascinating chaotic behavior. It alternately enters and exits the Kozai resonance: when outside of the Kozai resonance, it is at the same time in both the ν_5 and the ν_{16} secular resonances, σ_{16} librating around 180 degrees; when inside the Kozai resonance, the asteroid is no longer in ν_5 (recall that this resonance cannot occur for orbits with librating argument of perihelion) but it is still in the ν_{16} resonance, and σ_{16} librates around 0°. The transitions through the Kozai separatrix are repeatedly forced by the large oscillations of the inclination caused by the ν_{16} resonance. It is curious to know that the asteroid protagonist of this secular resonant dance has been named after James Williams, the pioneer investigator of asteroid secular dynamics.

The perturbation scheme explained in this section has also been used by Michel (1997) to study the dynamics in the secular resonances with the terrestrial planets' frequencies in the inner Solar System ($a < 2$ AU). Because the

ν_3 and ν_4 resonances – as well as the ν_{13} and ν_{14} resonances – may overlap, Michel computed a secular resonant normal form retaining the harmonics related to both resonances. Then, he numerically computed the Poincaré section of the resulting nonintegrable Hamiltonian in order to measure the extent of the chaotic zone. He found that the perihelion resonances can easily transport bodies from initially circular orbits to planet-crossing orbits and vice versa, while the nodal resonances can force changes of about 10 degrees in the inclination. Both phenomena are important to understand the evolution of near-Earth asteroids and their present orbital distribution.

Concerning the Kuiper belt, the amplitudes of the secular resonances ν_7, ν_8, ν_{17} and ν_{18} have been computed by Morbidelli *et al.* (1995a) through simplified models similar to Yoshikawa's and Nakai and Kinoshita's. Their results are in agreement with the numerically computed evolution and show that bodies on initially planar and circular orbits with semimajor axes in the ranges 35–36 AU and 40–42 AU are forced to cross the orbit of Neptune in a timescale of several 10^7 y. This behavior was first pointed out by Holman and Wisdom (1993) with numerical integration.

8.4.2 The anomalous case of the ν_6 resonance

We now separately discuss the case of the ν_6 resonance, which is, for its dynamical properties, by far the most important secular resonance in the asteroid belt.

Starting from Froeschlé and Scholl (1986), numerical simulations of real and fictitious asteroids have shown that the ν_6 resonance, unlike all other perihelion resonances, is able to pump the eccentricity of resonant bodies to values exceeding ~ 0.8. Farinella *et al.* (1994) showed for the first time that for some bodies in the ν_6 resonance the eccentricity increases up to unity. In these cases, the perihelion distance decreases to zero, so that the asteroids are forced to collide with the Sun. Figure 8.13 gives an example of this phenomenon: notice that the eccentricity of the body increases from 0 to 1 in an apparently regular manner, on a timescale of order 1 My. The large eccentricity attained by ν_6 resonant bodies, which makes them planet-crossers or Sun-grazers, explains why the location of the resonance corresponds to a region that is almost completely depleted of asteroids.

The ν_6 resonance has such a large effect on the eccentricity because it has an unusual dynamical structure, which is different from that of the typical pendulum-like models (4.4) obtained for all other resonances.

The reason for this difference can be understood in a qualitative way by looking at the top panel of Fig. 8.14, which reproduces the location of the ν_6 resonance on the (a, e) plane for different values of the inclination, as first

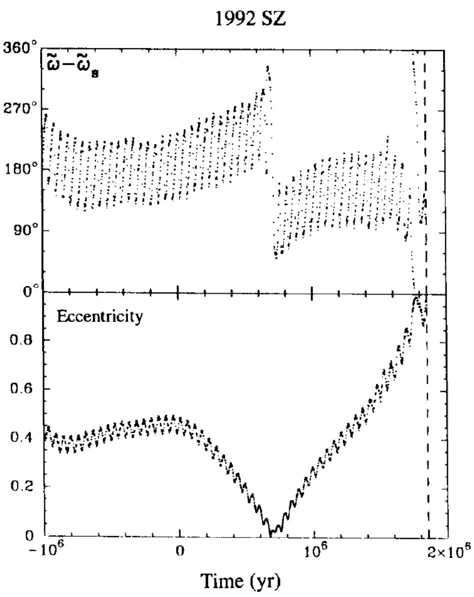
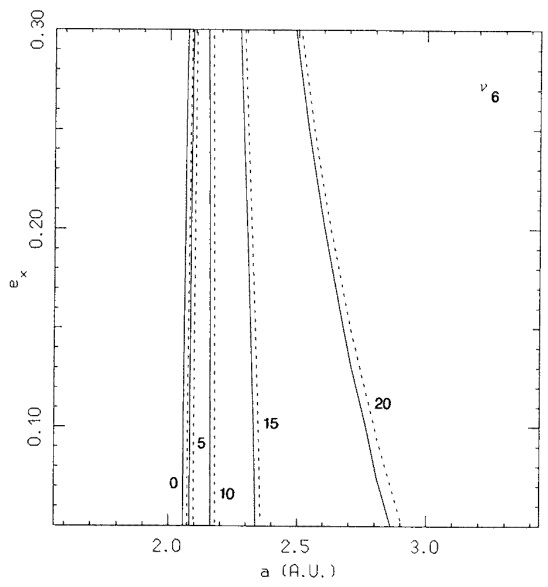


Figure 8.13: The evolution of the asteroid 1992 SZ in the ν_6 resonance, resulting from the numerical integration of the full equations of motion. Reprinted from Fig. 1 of Farinella *et al.* (1994) with permission from Nature, Macmillan Magazines Limited.

computed by Williams and Faulkner (1981). The striking feature is that, at least in the inner asteroid belt ($a < 2.5$ AU), the location of the ν_6 resonance is described by almost vertical lines; more specifically, for given inclination and semimajor axis, the resonance occurs for *all* values of the eccentricity. This situation is very atypical: for comparison, the bottom panel of Fig. 8.14 shows the equivalent plot for the location of the ν_5 resonance. Remember that a perihelion secular resonance pumps the eccentricity, keeping approximately constant the inclination; therefore, a body evolving in the ν_6 resonance, whose eccentricity increases under the resonant action, never leaves the exact resonant location, so that its eccentricity can grow indefinitely.

From the mathematical point of view, this situation is indicative of a loss of convexity of the resonant normal form Hamiltonian (8.38). We have already dicussed in Chapter 6 the role of convexity for the confinement of resonant motions, in the context of Nekhoroshev theory. More specifically, a nonconvex resonant normal form is a Hamiltonian of type (4.2) such that the coefficient β of its local expansion (4.3) is equal to zero. This implies that the frequency



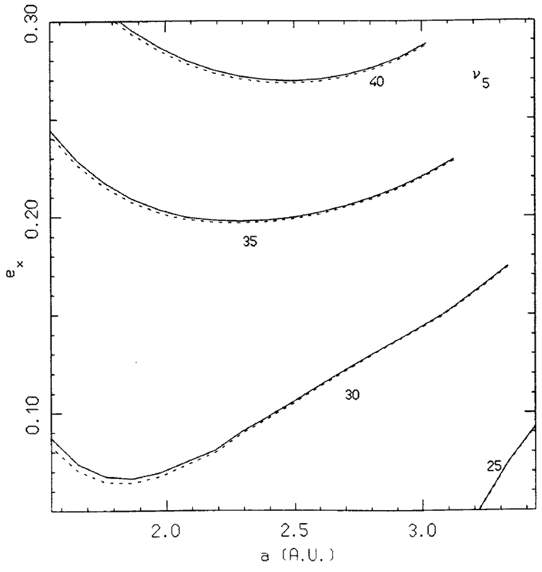


Figure 8.14: Top: the location of the ν_6 resonance with respect to proper semimajor axis and eccentricity, labeled by the corresponding values of the proper inclination. Bottom: the same, but for the ν_5 resonance. The solid curves reproduce the original computation by Williams and Faulkner (1981), while the dashed curves are drawn using more accurate values for the planetary proper frequencies g_6 and g_5. Reprinted from Figs. 5 and 6 of Morbidelli and Henrard (1991a), with permission from Kluwer Academic Publishers.

of the resonant angle is equal to zero for every value of its conjugate action (respectively φ_1 and $\hat{I}_1$ in 4.3). This is precisely what happens for the ν_6 resonance, where – for the appropriate values of semimajor axis and inclination – the frequency of the resonant angle $\sigma_6 = \psi_1 + \psi_2 - \varpi_6^*$ is equal to zero for every value of the eccentricity, i.e. of the conjugate action S, as indicated by the vertical lines in the top panel of Fig. 8.14. As a consequence of $\beta = 0$, in the approximation given by the local expansion (4.3) the action $\hat{I}_1$ would escape to infinity for most of the initial conditions (as can be easily seen by plotting the level curves of the Hamiltonian 4.4). However, when $\hat{I}_1$ increases too much, the local expansion (4.3) is no longer a good approximation of the real motion. The latter critically depends on the functional form, for large values of $\hat{I}_1$, both of $\mathcal{H}_0$ and of the coefficient of the resonant harmonic in the complete normal form (4.2). In other words, nonconvex resonant normal forms do not have a generic dynamical portrait, in contrast to convex resonant normal forms which all give pendulum-like motion. They therefore need to be studied from case to case.

Figure 8.15 gives an example of the dynamical structure of the ν_6 resonance. The various panels show the level curves of the secular resonant normal form Hamiltonian (8.38) for different values of the constant of motion C. The coordinates are $x = e_{\rm sp} \cos \sigma_6, y = e_{\rm sp} \sin \sigma_6$, equivalent to those used in Fig. 8.11 for the ν_5 resonance. The plots have been limited to the region $e_{\rm sp} \leq 0.65$, for technical difficulties in computing an accurate secular resonant normal form for larger values of the semiproper eccentricity. Panel b is drawn for a value of C corresponding to the fake asteroid with mean orbital elements $a = 2.3488$ AU, $e = 0.1802$, $i = 15.105°$, $\omega = 242.81°$, $\Omega = 138.750°$; the black dot denotes the present position of this fake asteroid in these x, y coordinates. According to the level curves, the semiproper eccentricity of this asteroid is expected to oscillate approximately between 0 and 0.5, while the resonant angle σ_6 librates around 180°, behavior confirmed by numerical simulation of the full equations of motion (see Morbidelli, 1993a). Notice that the level curves indicate the presence of an unstable equilibrium point at $x \sim -0.6, y = 0$ and of a separatrix, which surrounds the closed cycles centered on the stable equilibrium point (the latter located at $x \sim -0.27, y = 0$). Outside of the separatrix the level curves are open: every initial condition in this region leads to a semiproper eccentricity exceeding 0.65. From the shape of the level curves, it is reasonable to expect that the eccentricity may increase to much larger values than 0.65, possibly attaining the threshold $e = 1$. Actually, the evolution of the eccentricity shown in Fig. 8.13 is compatible with a diagram like Fig. 8.15b and with a body that initially evolves on a closed cycle near the separatrix (so that its eccentricity attains a first maximal value of ~ 0.4)

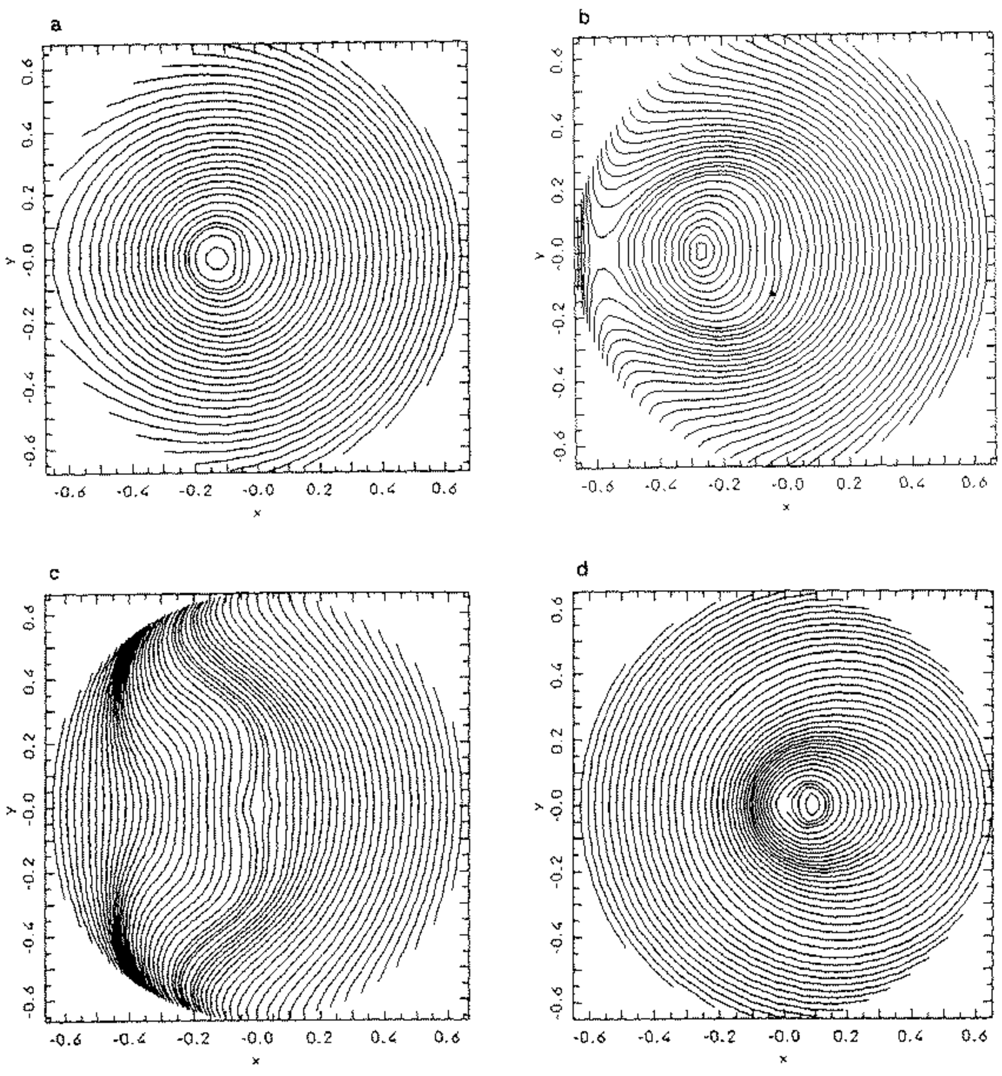


Figure 8.15: Dynamical portraits for the ν_6 resonance at $a \sim 2.35$ AU and different values of the inclination. See text for explanations. Reprinted from Fig. 12 of Morbidelli (1993a), with permission from Academic Press.

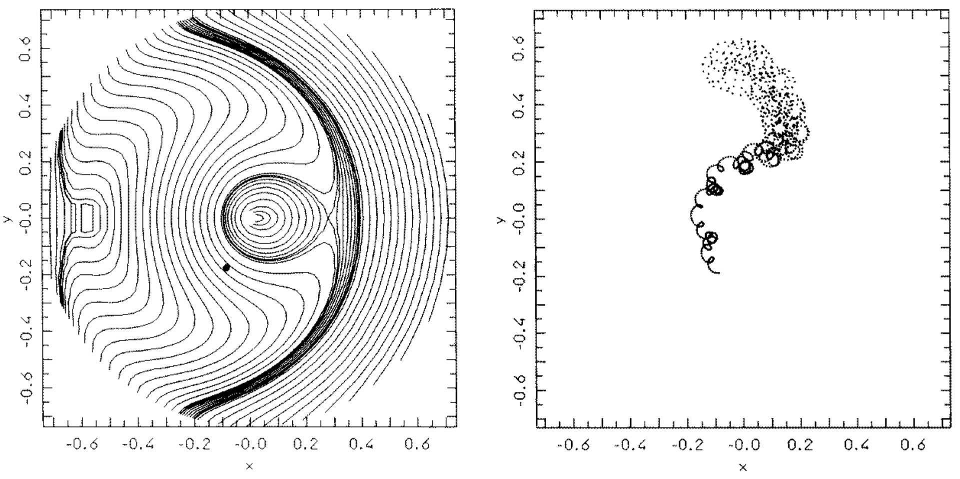


Figure 8.16: Left: dynamical portrait for the ν_6 resonance for a body with the same orbital elements as 6 Hebe, except the inclination, increased to 18.7°. The coordinates are $x = e_{\rm sp}\cos\sigma_g, y = e_{\rm sp}\sin\sigma_6$ and the black dot denotes the present position of the body. Right: the evolution of the same body as computed by numerical integration of the full equations of motion. The coordinates are $x = e\cos(\varpi - \varpi_6^*), y = e\sin(\varpi - \varpi_6^*)$, where e and ϖ are the osculating eccentricity and perihelion longitude of the body. Reprinted from Fig. 11 of Morbidelli (1993a), with permission from Academic Press.

and then crosses the separatrix joining one of the open level curves, so that its eccentricity first decreases to 0 and then increases to 1. The behavior of the resonant angle (top panel of Fig. 8.13) confirms this interpretation.[2] The fact that the body passes from one level curve to another, crossing the separatrix, should not surprise us. Remember from Chapter 4 that the single resonant normal form has a well-defined separatrix because it is integrable, but the remainder of the normal form generically breaks the integrability and splits the separatrix, thus creating a small chaotic zone; orbits in the chaotic zone do not follow exactly a level curve of resonant normal form, but alternate between the libration and the circulation regions.

Figure 8.15c is drawn for a value of C corresponding to the same asteroid of panel b, but with inclination increased to 15.6 degrees. The stable and

[2]Note that Fig. 8.13 shows the angle $\varpi - \varpi_{\rm S}$ instead of σ_6, where $\varpi_{\rm S}$ is the longitude of perihelion of Saturn. The angle $\varpi - \varpi_{\rm S}$ has large oscillations around σ_6 on a timescale of $\sim 5 \times 10^4$ y, because $\varpi_{\rm S} = \varpi_6^* + f(\varpi_k^*)$, where f is a function of all planetary secular perihelion frequencies (see formula 7.10).

unstable equilibrium points that characterized the dynamics in panel b have disappeared, and *all* level curves are open. This implies that, whatever the initial condition on this panel, the eccentricity is pumped by the resonance well beyond 0.65. Finally, panels a and d show the resonant portraits for values of C corresponding respectively to $i = 14.1°$ and $i = 17.1°$ of the considered asteroid. For these values of inclination, the asteroid is respectively significantly below and above the exact location of the secular resonance. The dynamics is characterized by the existence of only one stable equilibrium point, displaced from the center $x = 0, y = 0$ of the diagram. All orbits cycle around the stable equilibrium, the semiproper eccentricity periodically oscillating between a minimal and a maximal value. On most cycles, the resonant angle σ_6 circulates over the $[0, 360]$ degrees interval; its time derivative is positive in panel a and negative in panel d, as the frequency of the longitude of perihelion of an asteroid is respectively faster/slower than the planetary secular frequency g_6 for orbits below/above the ν_6 resonance. Decreasing the asteroid's inclination below the value corresponding to panel a or increasing it above the one corresponding to panel d (i.e. moving further from the resonance) the dynamical portrait would not change significantly with respect to those shown in these panels. In fact, there would always be a stable equilibrium point, slightly displaced from $e = 0$. This can be easily realized by taking into account the sole forcing terms with $k = 6$ in the linear equations of motion (8.6).

The diagrams shown in Fig. 8.15 cannot be considered as typical of the ν_6 resonance for every value of the semimajor axis. Because of the nonconvexity of the normal form, the phase portrait of the resonance may critically depend on the semimajor axis. As an example, Fig. 8.16 shows the level curves of the secular resonant normal form for an asteroid with mean elements $a = 2.43$, $e = 0.20$, $i = 18.7°$, $\omega = 239°$, $\Omega = 139°$. As one sees, the diagram looks very different from any of those in Fig. 8.15. This is not an artifact of the computation of the secular resonant normal form. In fact, the right panel of Fig. 8.16 shows the evolution of the asteroid computed in osculating coordinates by numerical integration of the full equations of motion, which well matches the level curve passing through the present position of the body in semiproper coordinates. As already discussed for the asteroid Barcelona, the level curve of the secular resonant normal form just traces the skeleton of the real secular dynamics: the evolution in osculating coordinates oscillates around the guiding level curve, because of all the nonresonant harmonics that have been averaged out in the construction of the secular resonant normal form.

Chapter 9

MEAN MOTION RESONANCES

9.1 A simple integrable approximation

This chapter is the first of a series devoted to mean motion resonances. Mean motion resonances constitute one of the most complicated aspects of the dynamics in the Solar System. They are different from the standard resonances described in Chapter 4, because Solar System dynamics is *degenerate*, namely it is characterized by the existence of both fast angles – associated with the orbital motions of the bodies – and slow angles – associated with the precessional motions of their orbits. In the study of the secular dynamics in Chapters 7 and 8, the degeneracy was removed by averaging over all fast angles. In the case of mean motion resonances this averaging is prevented by the resonant relationship, as discussed in Section 2.5. We therefore have to fully cope with degeneracy.

In this chapter we study the structure of the resonances between the mean motions of two bodies. We will concentrate on the case of an asteroid in resonance with one planet, because no two-body resonances are known among the planets of our Solar System. In the next chapter we will investigate the resonances among the mean motions of several bodies, and in Chapter 11 we will finally study the secular dynamics inside the mean motion resonances of low order.

A mean motion resonance (also often called *commensurability*) between an asteroid and the $\bar{\jmath}$-th planet occurs when $kn - k_{\bar{\jmath}} n_{\bar{\jmath}} \sim 0$, where k and $k_{\bar{\jmath}}$ are positive integers, $n_{\bar{\jmath}}$ is the mean motion frequency of the $\bar{\jmath}$-th planet and $n = 1/\Lambda^3$ is the mean motion frequency of the asteroid. The starting point for the study of a mean motion resonance is the normal form developed in

Section 2.5.2. The mean motion resonant normal form Hamiltonian reads:

$$\mathcal{H}_{\rm MMR} = \mathcal{H}_0(\Lambda,\Lambda_{\bar{\jmath}}) + \varepsilon\mathcal{H}_1(\Lambda,P,Q,p,q,k_{\bar{\jmath}}\lambda_{\bar{\jmath}}-k\lambda,e_j,\varpi_j,i_j,\Omega_j)\ . \qquad (9.1)$$

With respect to (2.45), for simplicity of notation, the terms of the normal form of higher order in ε have been included in $\varepsilon\mathcal{H}_1$ and all the superscripts that denote that the variables appearing in the normal form are the *semimean* modified Delaunay variables have been omitted. Conversely, we have emphasized that the normal form Hamiltonian depends on the *mean* eccentricities, perihelia, inclinations and nodes of all planets ($e_j,\varpi_j,i_j,\Omega_j$ with $j=1,\ldots,N$). With respect to the secular normal form (8.1), $\mathcal{H}_{\rm MMR}$ depends also on the mean longitudes of the asteroid and of the $\bar{\jmath}$-th planet $(\lambda,\lambda_{\bar{\jmath}})$, so that Λ and $\Lambda_{\bar{\jmath}}$ are not constant and the main term

$$\mathcal{H}_0(\Lambda,\Lambda_{\bar{\jmath}}) = -\frac{1}{2\Lambda^2} + n_{\bar{\jmath}}\Lambda_{\bar{\jmath}} \qquad (9.2)$$

cannot be dropped.

To study the dynamics of (9.1) we expand $\mathcal{H}_1$ in a power series of the planetary eccentricities and inclinations as in Section 8.2 and, for the moment, we consider only the leading term $\mathcal{H}_1^0$ of this expansion, which is independent of e_j and i_j. Because of the D'Alembert rules (see Section 1.9.3) its Fourier expansion has the form:

$$\mathcal{H}_1^0 = \sum_{m,s,r} c_{m,s,r}(\Lambda,P,Q)\exp\left[\iota(m(k_{\bar{\jmath}}\lambda_{\bar{\jmath}}-k\lambda)+sp+rq)\right] \qquad (9.3)$$

with m,r,s integer numbers such that $m(k_{\bar{\jmath}}-k)-s-r=0$. The Hamiltonian $\mathcal{H}_0+\varepsilon\mathcal{H}_1^0$ is not integrable, because (9.3) contains two independent combinations of the different angles, contrary to what happened in the secular problem (where $k=k_{\bar{\jmath}}=0$). Note however that the D'Alembert rules also imply that, for small inclinations i of the asteroid, the coefficients $c_{m,s,r}$ are proportional to $Q^{r/2}$, with $Q\sim i^2/2$. Therefore, if we restrict to the planar case $i=0$ the harmonics with $r\neq 0$ in (9.3) have null coefficients, so that the Hamiltonian $\mathcal{H}_0+\varepsilon\mathcal{H}_1^0$ is integrable as it depends only on the angle $k_{\bar{\jmath}}\lambda_{\bar{\jmath}}-k\lambda+(k_{\bar{\jmath}}-k)p$ and on its multiples. We therefore consider the Hamiltonian

$$\mathcal{H}_{\rm PC} = \mathcal{H}_0(\Lambda,\Lambda_{\bar{\jmath}}) + \varepsilon\mathcal{H}_1^0(\Lambda,P,k_{\bar{\jmath}}\lambda_{\bar{\jmath}}-k\lambda+(k_{\bar{\jmath}}-k)p;Q=0) \qquad (9.4)$$

as an integrable approximation of the resonant normal form (9.1). Notice that $\mathcal{H}_{\rm PC}$ is in fact the Hamiltonian of the mean motion resonant normal form in the framework of the so-called *planar circular restricted problem*, where the asteroid and the planets are assumed to move on the same plane, and the

orbits of the planets are circular. The planar circular restricted problem is only a quite crude approximation of the real dynamics, as we will see in this and in the following chapters, but nevertheless it is the suitable integrable approximation for the development of a perturbation study of the real mean motion resonant dynamics.

The D'Alembert rules also ensure that, for small eccentricity e of the asteroid, the coefficient of the leading harmonic $\exp \iota[k_{\jmath}\lambda_{\jmath} - k\lambda + (k_{\jmath} - k)p]$ is proportional to $P^{|k_{\jmath}-k|/2}$, with $P \sim e^2/2$. Therefore, mean motion resonances of type $kn - k_{\jmath}n_{\jmath} = 0$ have leading harmonics whose coefficients decay exponentially with $|k_{\jmath} - k|$. For this reason, astronomers call $|k_{\jmath} - k|$ the *order* of the resonance. This generates confusion with the *order* of the harmonic defined by mathematicians (related to the general decay of the coefficients in Fourier series, and adopted throughout this book), which in this case would be $k + k_{\jmath}$. To eliminate the ambiguity, we will hereafter refer to $|k_{\jmath} - k|$ as the *eccentricity order* of the resonance. In reality, both astronomers and mathematicians are correct in their motivations, as the coefficients of the leading harmonics of mean motion resonances decay as a quantity proportional to $\exp(-|k_{\jmath} - k|) \exp[-(k + k_{\jmath})]$.

To study the dynamics of Hamiltonian (9.4) – and, later, of its perturbation (9.1) – we first introduce the following set of canonical action–angle variables:

$$\begin{aligned}
S &= P \ , & \sigma &= \tfrac{k_{\jmath}\lambda_{\jmath} - k\lambda + (k_{\jmath}-k)p}{(k_{\jmath}-k)} \\
N &= \tfrac{k_{\jmath}-k}{k}\Lambda + P + Q \ , & \nu &= \tfrac{-k_{\jmath}\lambda_{\jmath} + k\lambda}{(k_{\jmath}-k)} \\
S_z &= Q \ , & \sigma_z &= \tfrac{k_{\jmath}\lambda_{\jmath} - k\lambda + (k_{\jmath}-k)q}{(k_{\jmath}-k)} \\
\tilde{\Lambda}_{\jmath} &= \Lambda_{\jmath} + \tfrac{k_{\jmath}}{k}\Lambda \ , & \tilde{\lambda}_{\jmath} &= \lambda_{\jmath} \ .
\end{aligned} \tag{9.5}$$

It is easy to check, using the Poisson bracket criterion (see Section 1.6), that the above transformation is canonical. The angle σ is called the *critical angle* of the mean motion resonance. The reason for which σ is not simply defined as $k_{\jmath}\lambda_{\jmath} - k\lambda + (k_{\jmath} - k)p$ is profound: the coefficients of the resonant harmonics $\exp \iota m[k_{\jmath}\lambda_{\jmath} - k\lambda + (k_{\jmath} - k)p]$ are proportional to $S^{m|k_{\jmath}-k|/2} \sim e^{m|k_{\jmath}-k|}$, for small S. If σ is defined as in (9.5), the resonant harmonics are $\exp \iota[m(k_{\jmath} - k)\sigma]$. As a consequence, the Hamiltonian has a property analogous to that given by the fourth D'Alembert rule (see Section 1.9.3), which allows the introduction of canonical Cartesian variables $x = \sqrt{2S}\sin\sigma, y = \sqrt{2S}\cos\sigma$ in order to remove the singularity at $S = e = 0$. In turn, the Hamiltonian is periodic in σ, with period $2\pi/|k_{\jmath} - k|$.

Note that the transformation (9.5) is singular for $k_{\jmath} = k$, i.e. for the 1/1 resonance, which occurs when the asteroid and the planet share the same orbit

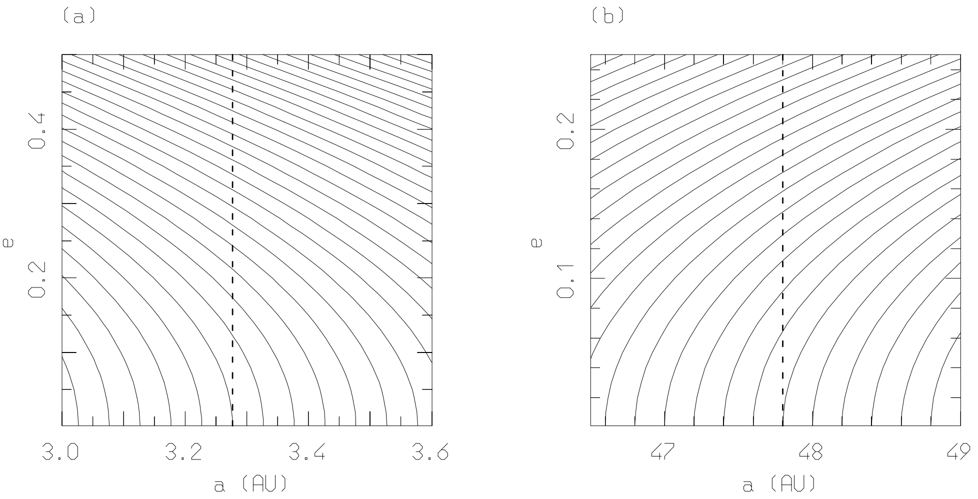


Figure 9.1: Level curves of N, as defined by (9.5) for the 2/1 resonance with Jupiter (panel a) and for the 1/2 resonance with Neptune (panel b). The values of N increase from left to right in panel (a) and from right to left in panel (b). The vertical dashed lines mark the unperturbed locations of these resonances.

(as for the Trojan asteroids). The 1/1 resonance therefore requires a specific formalism, which will be developed in Section 9.1.2

In the new variables the Hamiltonian (9.4) turns out to be:

$$\mathcal{H}_{\rm PC} = \mathcal{H}_0(\tilde\Lambda_{\jmath}, N, S, S_z) + \varepsilon \mathcal{H}_1^0(S, (k_{\jmath} - k)\sigma, N, S_z) \ , \qquad (9.6)$$

where, from (9.2),

$$\mathcal{H}_0 = -\frac{(k_{\jmath} - k)^2}{2k^2(N - S - S_z)^2} + n_{\jmath}\left[\tilde\Lambda_{\jmath} - \frac{k_{\jmath}}{(k_{\jmath} - k)}(N - S - S_z)\right] \ . \qquad (9.7)$$

It is then evident that $\tilde\Lambda_{\jmath}$, N and S_z are constants of motion for $\mathcal{H}_{\rm PC}$. Thus the term $n_{\jmath}\tilde\Lambda_{\jmath}$ can be dropped from (9.6). Remember that S_z (i.e. Q) has been set equal to 0 to obtain the Hamiltonian of the planar problem; therefore, the interesting part of the dynamics concerns only the variables S and σ and depends parametrically on the value of N. We can therefore represent the dynamical structure of the resonance by plotting the level curves of the Hamiltonian (9.6) on the S, σ plane, for different values of N.

At this point, it is useful to keep in mind the relationship between the actions S, N and the orbital elements a, e. The relationship $N =$constant defines a curve in the semimajor axis vs eccentricity plane. Figure 9.1 shows

these curves in the case of *interior resonances* ($k_{\bar{\jmath}} > k$; left panel) and in the case of *exterior resonances* ($k_{\bar{\jmath}} < k$; right panel). Therefore, the motion of S on a curve $N =$constant produces coupled oscillations of a and e. In the unperturbed problem ($\varepsilon = 0$), the exact resonance occurs for those values of S such that $\dot{\sigma} = \partial\mathcal{H}_0/\partial S = 0$, namely for:

$$S = N - \frac{k_{\bar{\jmath}} - k}{(k^2 k_{\bar{\jmath}} n_{\bar{\jmath}})^{1/3}} \ , \quad \text{i.e. } L = \sqrt{a} = \left(\frac{k}{k_{\bar{\jmath}}}\frac{1}{n_{\bar{\jmath}}}\right)^{1/3} \ , \ \forall e \ . \tag{9.8}$$

The unperturbed location of the resonance is represented by the vertical dashed line in Fig. 9.1.

The dynamical portrait of the mean motion resonances of first order in the eccentricity ($|k_{\bar{\jmath}} - k| = 1$) was computed long ago (Poincaré, 1902a, 1902b; Message, 1966; Schubart, 1964, 1968; Henrard and Lemaître, 1983). Figure 9.2 shows how the typical dynamical portrait of a resonance with $k_{\bar{\jmath}} = k + 1$ changes, as a function of N. If N is small enough, the corresponding curve $N =$constant in Fig. 9.1 does not cross the line denoting the unperturbed location of the resonance. The resulting dynamics on the S, σ plane presents only one stable equilibrium point at $\sigma = 0$ close to the coordinate center $S = 0$ (the distance from the center decreases with N); no unstable equilibria and no separatrices are visible (top panels). The situation changes when N is equal to a threshold value N_* (middle left panel). One of the level curves shows an angular point at $\sigma = \pi$; the dynamics being regular (there are no singularities in the equations of motion) the angular point must be an unstable equilibrium point, and the level curve passing through this point is the *separatrix* connecting the unstable equilibrium to itself (see Chapter 4). For $N > N_*$, the dynamical portrait is again different (middle right panel). Now a third equilibrium point appears, also at $\sigma = \pi$ but stable, and the separatrix originating from the unstable point now presents two loops. This is a typical portrait of a resonance when the action–angle variables are used as polar coordinates, as shown in Fig. 4.1. With increasing value of N beyond the threshold N_*, the stable equilibrium at $\sigma = 0$ and the unstable equilibrium migrate to larger values of S, while the stable equilibrium at $\sigma = \pi$ approaches the center $S = 0$ (bottom panels). As explained in Section 4.1, the reason for which the coordinate center $S = 0$ is never an equilibrium point is that the coefficient of the leading resonant harmonic $\exp \iota\sigma$ is proportional to $\sqrt{S}$. Because of this specific property, for the resonances of first eccentricity order we need to distinguish between the resonant region and the libration regions. As sketched in Fig. 9.3, the resonant region is – properly speaking – that enclosed by the separatrices that surround the stable equilibrium at $\sigma = 0$ (the shaded area in the top panels). When the separatrices do not exist (as

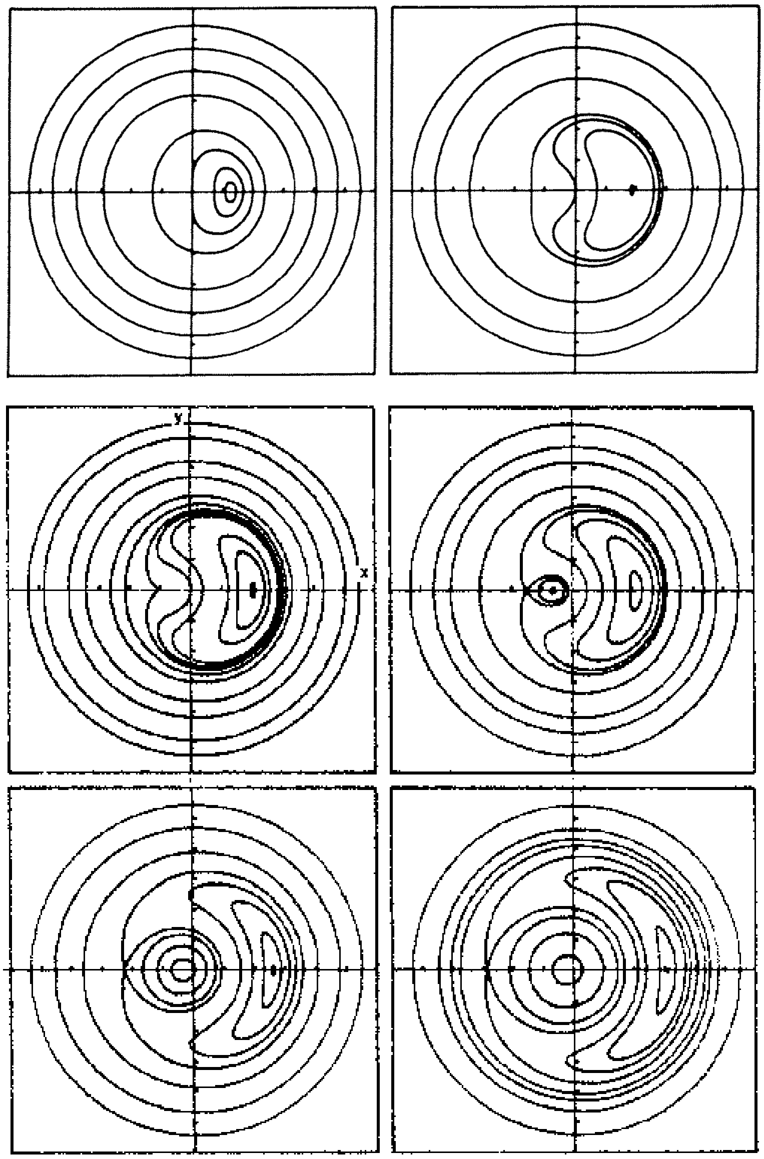


Figure 9.2: Phase portraits of an interior resonance of first order in the eccentricity, for different values of N. The coordinates are $\sqrt{2S}\cos\sigma \sim e\cos\sigma$ on the horizontal axis and $\sqrt{2S}\sin\sigma \sim e\sin\sigma$ on the vertical axis. See text for comments. Reprinted from Figs. 6 and 7 of Henrard and Lemaître (1983), with permission from Kluwer Academic Publishers.

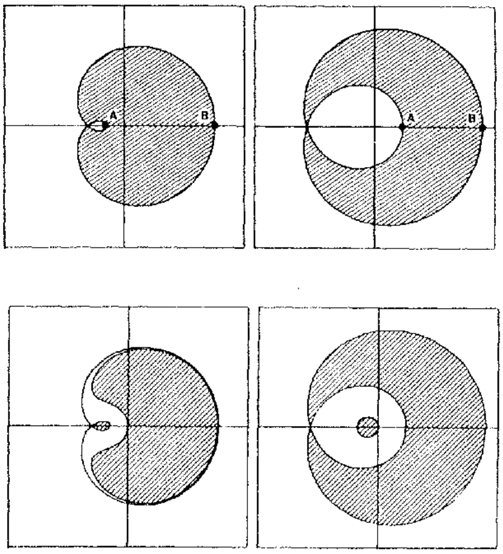


Figure 9.3: Distinction between resonant regions and libration regions for resonances of first order in the eccentricity, for two values of N. The shaded areas show the resonant regions on the top panels, and the libration regions in the bottom panels. See text for the definition of these regions. Reprinted from Figs. 2 and 3 of Henrard and Lemaître (1983), with permission from Kluwer Academic Publishers.

in the top panels of Fig. 9.2), the resonant region does not exist. Conversely, the libration regions are the sets of orbits on which σ librates around either 0 or π (the shaded areas in the bottom panels). As one sees, in some cases resonant orbits may have a circulating σ, while orbits with librating σ can be nonresonant. In this case, in fact, the geometrical concept of libration does not have a real dynamical significance. Historically, the orbits with σ librating around π are called *apocentric librators*, because the stable equilibrium at $\sigma = \pi$ corresponds to a periodic orbit on which the asteroid is at aphelion every time that it is in conjunction with the resonant planet ($p = \lambda + \pi$ when $\lambda = \lambda_{\bar{\jmath}}$). Despite their libration, these orbits are dynamically equivalent to circulating orbits in a pendulum.

The portrait of the dynamics of exterior resonances of first order in the eccentricity ($k = k_{\bar{\jmath}} + 1$) is equivalent to that of Fig. 9.2, but with σ rotated by 180 degrees. The 1/2 resonance ($k = 2$, $k_{\bar{\jmath}} = 1$) constitutes the only exception. In the latter, for N larger than some threshold, inside the resonant region the equilibrium point at $\sigma = \pi$ becomes unstable, and two new stable equilibria

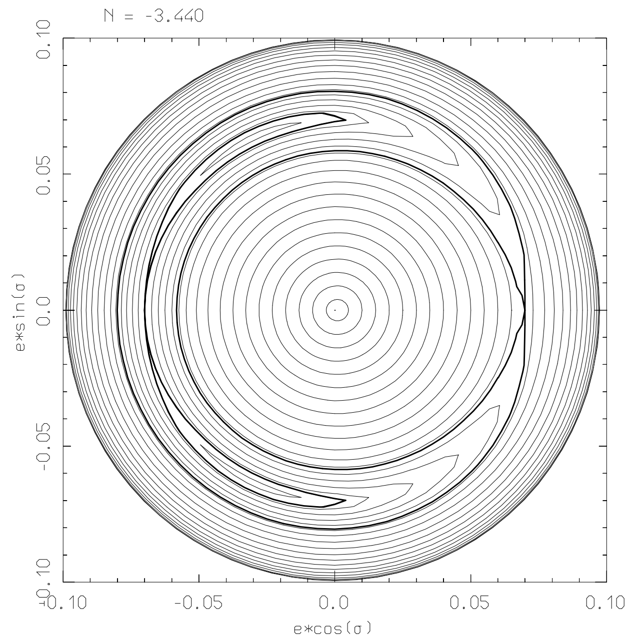


Figure 9.4: Phase portrait of the 1/2 resonance with Neptune, showing islands of *asymmetric libration*. The bold curves show the separatrices that originate from the unstable equilibria at $\sigma = 0$ and $\sigma = \pi$. Adapted from Thomas (1998).

appear in a symmetric position with respect to the horizontal axis (Fig. 9.4). A separatrix, connecting the unstable equilibrium at $\sigma = \pi$ to itself, surrounds the two stable equilibria. As a consequence, in the resonant region we distinguish between *symmetric* and *asymmetric* librators: the symmetric librators are the orbits with σ librating around π, in the region bounded by the separatrix generated by the unstable equilibrium point at $\sigma = 0$; the asymmetric librators have σ librating around some different value, and are in one of the two islands bounded by the separatrix generated by the unstable equilibrium point at $\sigma = \pi$. The reason for the appearance of the asymmetric libration regions is the following. Because of the specific values of the coefficients c_1 and c_2 at the location of the 1/2 resonance, the harmonic $c_2 S \exp \iota 2\sigma$ in the perturbation $\mathcal{H}_1^0$ starts to dominate over the "leading" harmonic $c_1 \sqrt{S} \exp \iota\sigma$, for S large enough. Remember from Fig. 9.1 that, at the resonance location, the value of S increases with N; as a consequence the asymmetric librations must appear for N sufficiently large. The existence of asymmetric librations was first noted by Message (1958) and then further investigated by Schubart (1964) and Beaugé (1994) and we refer to these papers for further details.

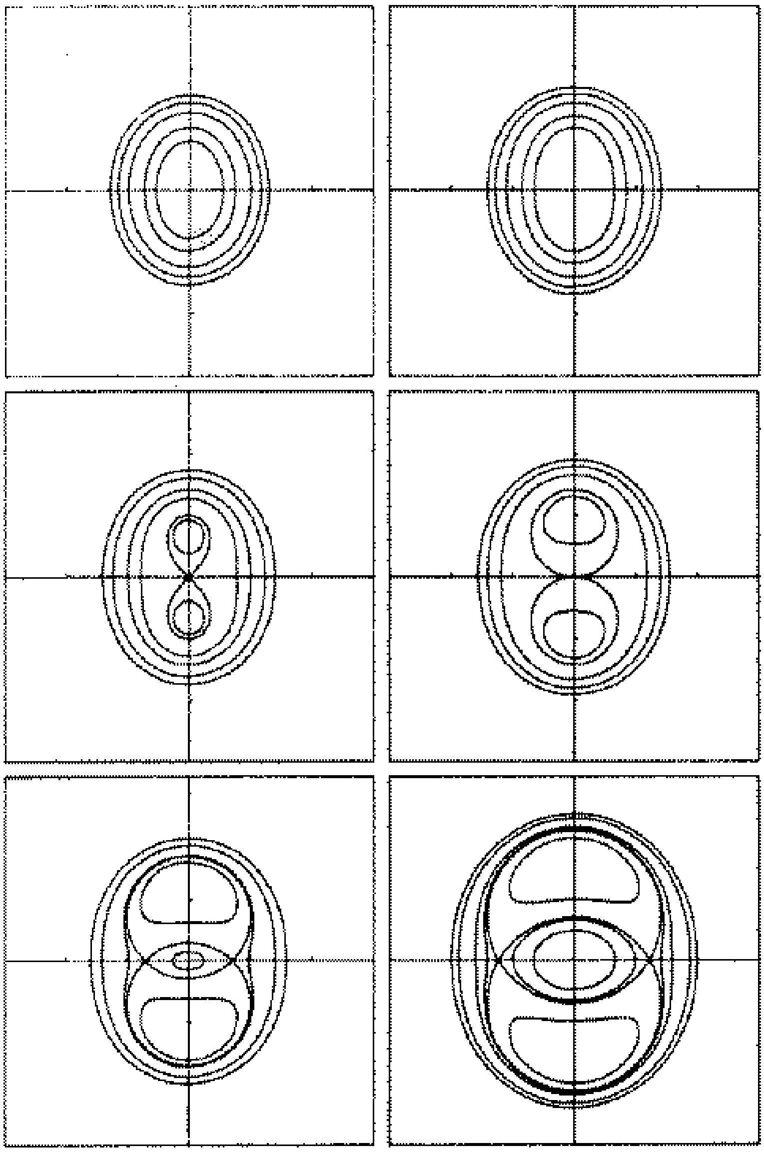


Figure 9.5: Phase portraits of a resonance of second order in the eccentricity, for different values of N. The coordinates are $\sqrt{2S}\cos\sigma \sim e\cos\sigma$ on the horizontal axis and $\sqrt{2S}\sin\sigma \sim e\sin\sigma$ on the vertical axis. See text for comments. Reprinted from Fig. 1 of Lemaître (1984), with permission from Kluwer Academic Publishers.

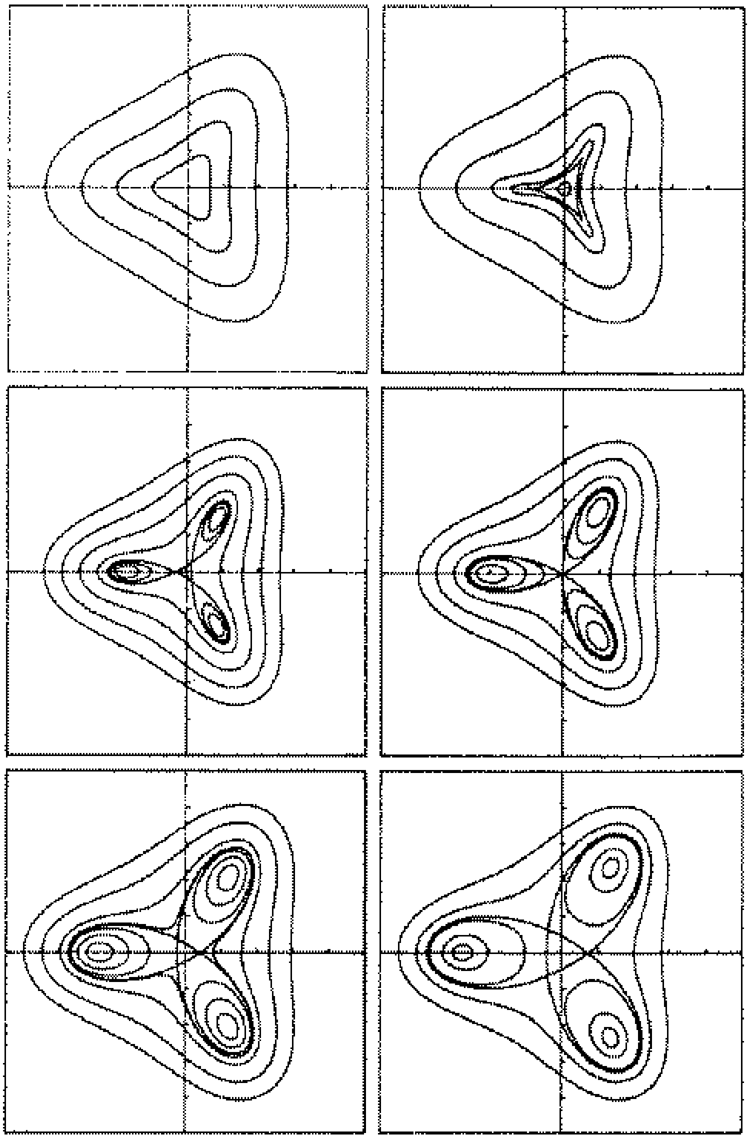


Figure 9.6: Phase portraits of an exterior resonance of third order in the eccentricity, for different values of N. The coordinates are $\sqrt{2S}\cos\sigma \sim e\cos\sigma$ on the horizontal axis and $\sqrt{2S}\sin\sigma \sim e\sin\sigma$ on the vertical axis. See text for comments. Reprinted from Fig. 4 of Lemaître (1984), with permission from Kluwer Academic Publishers.

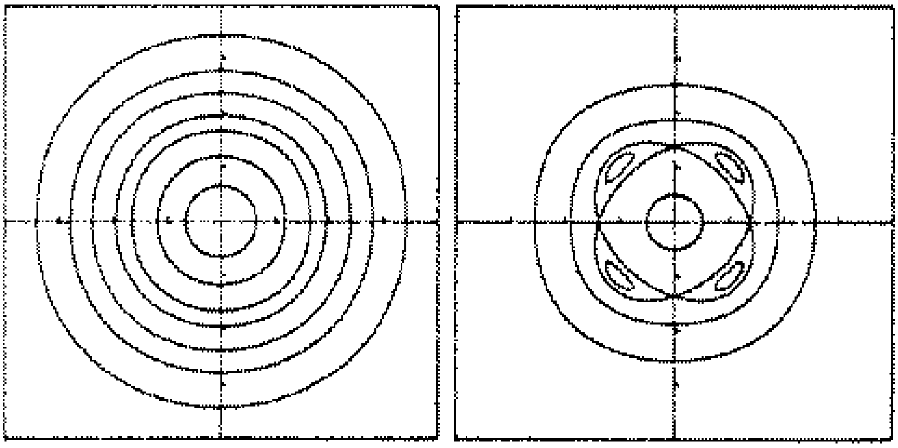


Figure 9.7: Phase portraits of a resonance of fourth order in the eccentricity, for different values of N. The coordinates are $\sqrt{2S}\cos\sigma \sim e\cos\sigma$ on the horizontal axis and $\sqrt{2S}\sin\sigma \sim e\sin\sigma$ on the vertical axis. See text for comments. Reprinted from Fig. 7 of Lemaître (1984), with permission from Kluwer Academic Publishers.

The dynamical portraits of resonances of eccentricity order larger than one have been illustrated by Lemaître (1984). Figures 9.5 refers to interior resonances of order 2 in the eccentricity ($k_{\bar{\jmath}} = k + 2$). For N smaller than some threshold, the coordinate center $S = 0$ is a stable equilibrium point around which all other orbits circulate (top panels). Increasing N a first bifurcation occurs when the center $S = 0$ becomes unstable and two stable equilibria appear at 90 and 270 degrees (middle panels). A separatrix connecting the origin to itself surrounds the stable equilibria, delimiting two resonant regions, where σ librates. For N larger than a second threshold value, the dynamical portrait changes again. The center $S = 0$ is stable, while two unstable equilibria exist at $\sigma = 0, \pi$. The separatrix now originates from these equilibria; the resonant regions are still those where the orbits librate around 90 or 270 degrees. Exterior resonances have exactly the same dynamical portrait, with the exception of the 1/3 resonance, where the equilibria at 90 and 270 degrees become unstable for N larger than some threshold, and two new stable equilibria appear in a symmetric position with respect to the vertical axis, inside each resonant region (see Beaugé, 1994). As for the 1/2 resonance, we call *asymmetric librators* the orbits that librate around one of these stable equilibria.

Figure 9.6 refers to exterior resonances of order 3 in the eccentricity ($k = k_\jmath + 3$), with the exception of the 1/4 resonance, which again exhibits asymmetric librations. The dynamical portrait of interior resonances of order 3 in the eccentricity is like that of Fig. 9.6, but with σ rotated by 180 degrees.

Finally, Fig. 9.7 shows the dynamical portrait of resonances of order 4 in the eccentricity. Interior and exterior resonances share the same portrait, with the exception of the 1/5 resonance which has asymmetric librators.

9.1.1 Phase protection from planetary collisions

In the previous section, we have seen that in interior resonances the unstable equilibria are located at $\sigma = \sigma_m = \pi[1 + 2mk_\jmath/(k_\jmath - k)]$, where m is an integer in the range $[0, k_\jmath - k)$. Note that, when $\sigma = \sigma_m$, one has $p = \lambda + \pi$ whenever $\lambda_\jmath = \lambda + 2m\pi$. In words, on the unstable equilibria the asteroid is at *aphelion* in correspondence to one of the $k_\jmath - k$ different conjunctions with the resonant planet. Conversely, in exterior resonances unstable equilibria are located at $\sigma = \sigma_m = 2mk_\jmath\pi/(k_\jmath - k)$, and correspond to orbits on which the asteroid is at *perihelion* when one of the conjunctions with the resonant planet occurs. Therefore, unstable equilibria always correspond to the resonant orbits that give the closest possible approach between the asteroid and the resonant planet. Because of their libration, resonant orbits avoid this configuration of closest approach.

This property is important when the eccentricity is large, because in this case the closest approach may correspond to a physical collision with the planet. In fact, when the value of N is such that the eccentricity of the unstable equilibria is equal to $a_\jmath/a - 1$ for interior resonances, or $1 - a_\jmath/a$ for exterior resonances (a and $a_\jmath$ are the semimajor axes of the asteroid and of the resonant planet respectively) the unstable equilibria correspond to orbits that collide at aphelion (or perihelion) with the resonant planet. When N is larger than this value, the dynamical portrait of the resonance typically looks like the one shown in Fig. 9.8, computed for the 5/6 resonance with Neptune. The bold curve denotes the set of points e, σ that correspond to a collision with the planet. This curve is computed by setting $r = a_\jmath$ and $f + \varpi = \lambda_\jmath$, where r and f are the heliocentric distance and the true anomaly of the asteroid, given in (1.5), which can be easily expressed as functions of N, S and σ. All orbits that cross this curve will undergo collisions or very close encounters with the planet. Note that the resonant orbits, because of their libration around $\sigma = \pi$, all avoid intersecting the collisional curve. The mean motion resonance therefore provides to the resonant orbits a phase protection from collisions. Pluto (the largest object known in the Kuiper belt) is the most famous example of this property. It is in 2/3 resonance with Neptune, and

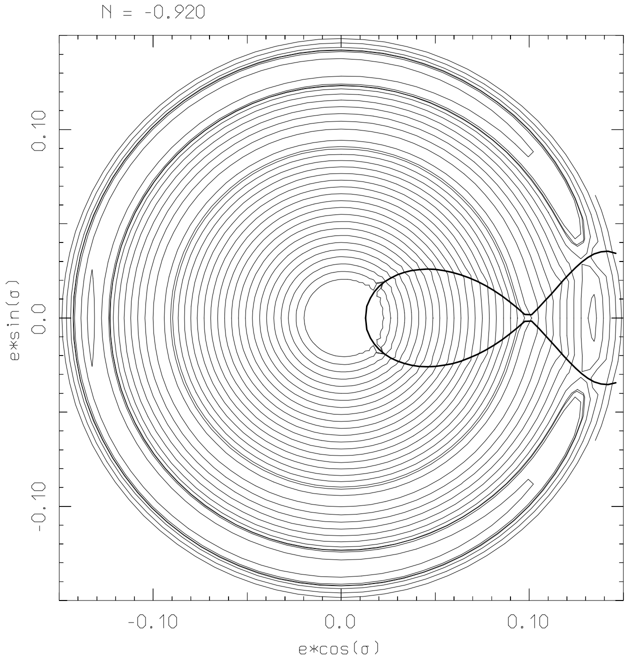


Figure 9.8: A phase portrait of the 5/6 resonance with Neptune; the bold curve denotes the set of points (e, σ) for which a collision with the planet can occur. Only the librational trajectories that do not intersect this curve may be dynamically stable. Adapted from Thomas (1998).

penetrates inside Neptune's orbit. However the libration of σ never allows it to cross the collisional curve, so that every time that Pluto crosses Neptune's orbit, it is always far from Neptune's position. This kind of ballet between Pluto and Neptune is believed to have been stably ongoing since the primordial phases of the Solar System.

In Fig. 9.8 one can also notice that the equilibrium point at $\sigma = 0°$, usually unstable for exterior resonances of first order in the eccentricity, is now stable. A small region of stability surrounds this point, bounded by the collisional curve. For increasing values of N (i.e. for larger eccentricity of resonant orbits) the region of stability around the equilibrium at $\sigma = 0$ increases in size, while that around the equilibrium at $\sigma = \pi$ shrinks. It should be finally remarked that a mean motion resonance provides protection only from close encounters with the resonant planet; as a consequence, all orbits must eventually become unstable if the eccentricity is large enough to intersect nonresonant planets' orbits.

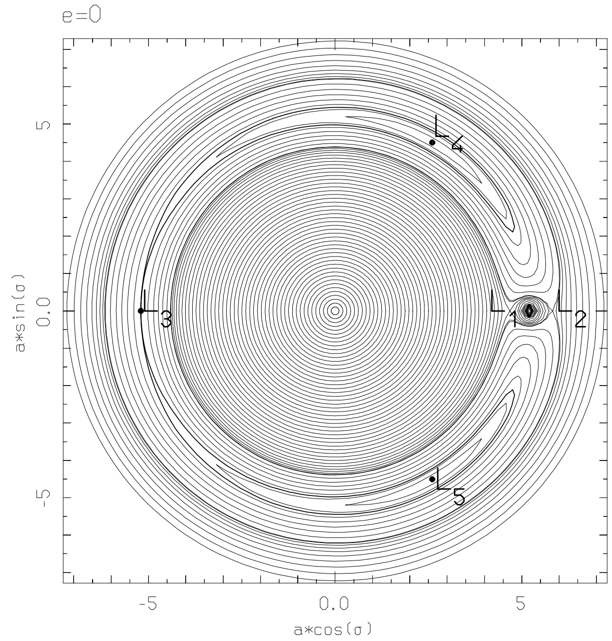


Figure 9.9: Phase portrait of the 1/1 resonance with Jupiter for a value of N corresponding to $e = 0$.

9.1.2 The case of the 1/1 resonance

In order to study a 1/1 resonance with a planet, we need to modify the choice of the resonant variables done in (9.5), which are singular when $k_{\bar{\jmath}} = k$. Suitable canonical variables for this resonance are the following:

$$\begin{aligned} S &= \Lambda \ , & \sigma &= \lambda - \lambda_{\bar{\jmath}} \\ N &= P \ , & \nu &= p \\ S_z &= Q \ , & \sigma_z &= q \\ \tilde{\Lambda}_{\bar{\jmath}} &= \Lambda_{\bar{\jmath}} + \Lambda \ , & \tilde{\lambda}_{\bar{\jmath}} &= \lambda_{\bar{\jmath}} \ . \end{aligned} \tag{9.9}$$

Once these variables are introduced, the study of the 1/1 resonance's dynamics can be done as explained above for the other resonances.

Figure 9.9 shows the portrait of the 1/1 resonance with Jupiter for $N = 0$ (i.e. $e = 0$); for a better view of the dynamics, we use $a = S^2$, instead of S, as the radial coordinate. There are two points where the Hamiltonian has a polar singularity: the center of the coordinates $a = 0$, and the point at $\sigma = 0, a = 5.2$. The first one corresponds to the location of the Sun; the quasi-circular trajectories around it represent quasi-keplerian heliocentric orbits, whose semimajor axis is only slightly perturbed by the resonant planet

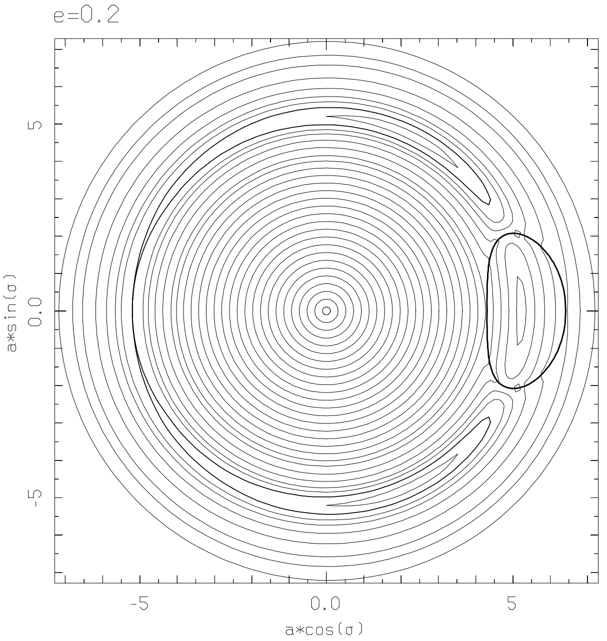


Figure 9.10: Phase portrait of the 1/1 resonance with Jupiter for a value of N corresponding to $e = 0.2$ at $a = 5.2$ AU.

as a function of σ. The second point corresponds to the position of Jupiter, which is fixed in these coordinates. In addition to these two points, five equilibrium points are visible in the figure: the well-known *Lagrangian points.* The two denoted by L_4 and L_5 are stable, and form two equilateral triangles with the positions of the Sun and the planet. The librations around these points are usually called *tadpole librations*, and are equivalent to the asymmetric librations existing in the exterior mean motion resonances of type $1/m$ (i.e. $k_{\jmath} = 1, k \in \mathbf{N}$; see for instance Fig. 9.4). These librations are delimited by a separatrix connecting the unstable equilibrium L_3 to itself. The Greek and Trojan asteroids follow tadpole librations. Also visible in Fig. 9.9 is a large region of libration around $\sigma = \pi$, bounded by the separatrices that originate at the two other unstable equilibria, L_1 and L_2. The orbits in this region are usually called *horseshoe librators*, and are equivalent to the symmetric librators discussed for Fig. 9.4.

The separatrices that originate at L_1 do not coincide with those that originate at L_2; but because they are very close to each other, this is hard to see in Fig. 9.9. On the other hand, these separatrices cannot mutually intersect, because the mean motion resonant normal form for the planar circular problem is integrable.

The separatrix connecting L_1 to itself bounds a small region of libration around $a = 5.2, \sigma = 0$. These trajectories orbit around the planet, in a satellite-type motion. Because $\dot{\sigma}$ must be positive when $a \ll a_{\jmath}$, the trajectories at the center of Fig. 9.9 must circulate around the origin in a counterclockwise motion. Then, by continuity, the horseshoe librations must be clockwise, and the satellite-type trajectories around Jupiter must be counterclockwise. Jupiter's rotation around its spin axis is also counterclockwise, so that the satellite-type trajectories are said to be *prograde*. The mean distance of L_1 and L_2 from the planet is called the *Hill's radius* and is equal to $a_{\jmath}(\varepsilon_{\jmath}/3)^{1/3}$, where $\varepsilon_{\jmath}$ is the mass (in solar units) of the planet.

Figure 9.10 shows the resonant portrait of the 1/1 resonance for a value of N corresponding to orbits with $e = 0.2$ at 5.2 AU. We again recognize two stable equilibria (now slightly offset with respect to $\sigma = \pm 60°$), centers of tadpole librations. The latter are bounded by a separatrix that originate at the unstable equlibrium at $a \sim 5.2$ AU, $\sigma = 180°$. The new feature with respect to Fig. 9.9 is the bold curve surrounding the point $a = 5.2$ AU, $\sigma = 0°$, which denotes the set of a, σ for which a collision with the planet can occur. This curve has been computed as explained in Section 9.1.1, and is analogous to that shown in Fig. 9.8. Because for this value of N the collision curve bounds a region of radius larger than the Hill's radius, the unstable equlibria denoted by L_1 and L_2 in Fig. 9.9 have disappeared, as well as their associated separatrices.[1] As a consequence, the maximal amplitude of horseshoe librations is now dictated by the position of the collision curve. Well inside the collision curve, there are again trajectories that librate around the position of Jupiter ($a = 5.2, \sigma = 0$) in a stable satellite-type motion. By continuity with the direction of motion of σ on the orbits that cross the collision curve ($\dot{\sigma} > 0$ on the orbits with $a < 5.2$ AU and $\dot{\sigma} < 0$ on those with $a > 5.2$ AU), the librations around Jupiter are now clockwise, i.e. *retrograde* with respect to Jupiter's spin. Retrograde satellites are therefore characterized by large *heliocentric* eccentricity,[2] while prograde satellites may have only small heliocentric eccentricity.

9.2 Mean motion resonance overlapping

In Chapter 6 we have seen the importance of the concept of resonance overlapping to understand the global stability or instability of a given dynamical

[1] For a value of N corresponding to a more moderate eccentricity, the radius of the region bounded by the collision curve would be smaller than the Hill's radius, and the two unstable equilibria at $\sigma = 0°$ would persist, corresponding to the so-called *Lyapunov orbits*.

[2] Which should not be confused with the *planetocentric* eccentricity.

system.

To check the possible overlapping of mean motion resonances we need to compute the width of each resonance in the semimajor axis vs. the eccentricity plane. This is done as follows. As shown in Section 9.1, on each surface $N=$constant, the separatrices (when they exist) typically intersect the axis $\sigma = \sigma_{\rm stab}$ twice, where $\sigma_{\rm stab}$ is the value of σ of the stable equilibrium point, at the center of the resonant region. Denote by $S_1(N)$ and $S_2(N)$ the values of S of these two intersection points (i.e. the minimal and maximal values of S attained on the separatrices). Then, compute the images $(a_1(N), e_1(N))$ and $(a_2(N), e_2(N))$ of these two points on the (a, e) plane, by inversion of (9.5) and (1.69), namely by using the formulæ:

$$a = \left[\frac{k}{k_{\jmath} - k}(N - S)\right]^2 \quad , \quad e = \left[1 - \left[1 - \frac{k_{\jmath} - k}{k}\frac{S}{N - S}\right]^2\right]^{1/2} \quad . \tag{9.10}$$

Finally, connect all the points $(a_1(N), e_1(N))$ obtained for increasing values of N; do the same for the points (a_2, e_2). The two curves so obtained represent the sections of the separatrices of the resonance at $\sigma = \sigma_{\rm stab}$. Equivalently, denote by $S_{\rm stab}(N)$ the value of S for the stable equilibrium, and compute its image $(a_{\rm stab}(N), e_{\rm stab}(N))$, thus obtaining a curve representing the family of the stable equilibrium points, parametrized by N. Figure 9.11 shows these curves for the 2/1 and for the 3/1 resonances with Jupiter in the asteroid belt. Also plotted are some curves $N=$constant. As we have seen in Section 9.1, in the integrable approximation of the planar circular problem each resonant orbit shows a coupled oscillation of a and e on a curve $N=$constant; the two extremes of this oscillation, achieved when $\sigma = \sigma_{\rm stab}$, lie symmetrically on each side of the family of stable equilibria and within the separatrices. The larger the amplitude of libration of σ, the closer to the separatrices are the two extremes of such oscillation. Notice that in both cases the width of the resonance increases with the eccentricity; this is because the coefficients of the resonant harmonics are proportional to some positive power of e.

In the case of the 2/1 resonance, one of the two separatrices does not reach $e = 0$ (Fig. 9.11a). This is because in mean motion resonances of first order in the eccentricity, when the separatrix first appears (see the middle-left panel of Fig. 9.2), it already crosses the axis $\sigma = \sigma_{\rm stab}$ at $e > 0$. The second feature of these resonances is that, along the family of stable equilibria, $a \to 0$ when $e \to 0$ (for exterior resonances $a \to +\infty$ when $e \to 0$); conversely, for resonances of larger eccentricity order, the family of stable equilibria hits the $e = 0$ axis for some definite value of a (Fig. 9.11b).

The recipe described above for the determination of the resonance's width cannot be applied for those resonances and those values of N that exhibit

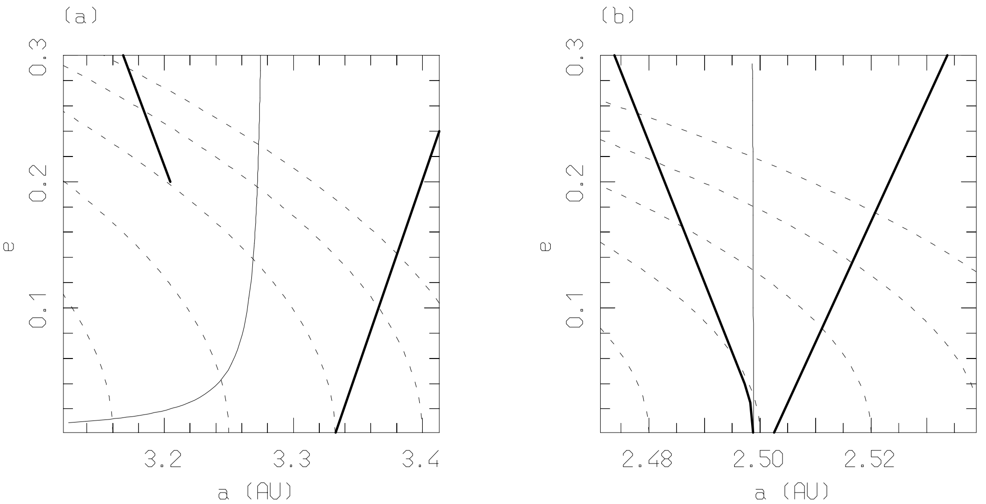


Figure 9.11: Separatrices (bold lines) and families of stable equilibria (solid light lines) for the 2/1 resonance (panel a) and 3/1 resonance (panel b) with Jupiter. Some level curves of N are also shown dashed.

asymmetric librations, as well as for eccentricities above the planet-crossing curve. In the first case, the width of the resonance is defined by the separatrix that bounds the symmetric librations (see Fig. 9.4). The minimal and maximal values of S on this separatrix are not attained when $\sigma = \sigma_C$, the latter being the center value of symmetric librations. This makes the computation of $S_1(N)$ and $S_2(N)$ somewhat more complicated. However, for many practical purposes, one can approximate S_1 and S_2 by the values of S on the separatrix when $\sigma = \sigma_C$, the difference with the real values being generally small. In the second case, that is for those values of N such that the collision curve appears on the S, σ portrait and there is no well-defined separatrix (Fig. 9.8), the width of the resonance is defined by the trajectory of largest librational amplitude that does not cross the collision curve. S_1 and S_2 are the minimal and maximal values of S on this trajectory. Because the largest libration amplitude of the noncollisional resonant orbits decreases with increasing N, the widths of mean motion resonances appear to decrease with increasing eccentricity, above the planet-crossing curve $a_{\bar{J}} = a(1 \pm e)$.

The first systematic computation of the widths of mean motion resonances in the asteroid belt was done by Dermott and Murray (1983). They studied all the mean motion resonances with Jupiter of type $n/n_{\rm J} = (2m+1)/m, (2m-1)/m, (3m+1)/2m$ – where $n_{\rm J}$ is the mean motion of Jupiter and m is an integer – that are located between 2.4 and 4 AU, and the 8/3 resonance at

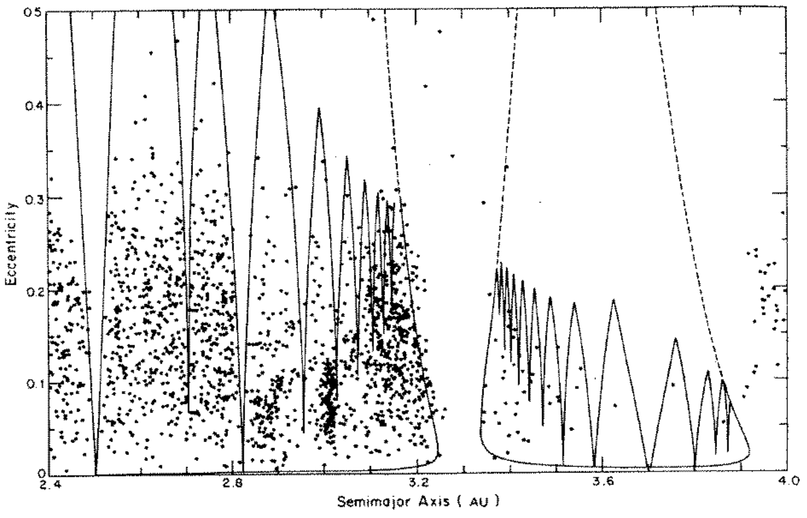


Figure 9.12: Distribution of the osculating elements of the asteroids known in 1983 with $2.4 < a < 4.0$ AU The solid lines represent the libration widths associated with the main mean motion resonances with Jupiter. In the region of resonance overlap, the libration widths of the 2/1 and 3/2 resonances are represented by dashed lines. Reprinted from Fig. 5 of Dermott and Murray (1983), with permission from Nature, Macmillan Magazines Limited.

2.7 AU (Fig. 9.12). These resonances are those of largest width in the considered semimajor axis range. With respect to the recipe explained above, a few simplifications have been made by the authors in order to make the computations easier. First, for the 2/1 and the 3/2 resonances (the only resonances of first order in the eccentricity among those studied) the width has been computed referring to the libration regions, rather than to the resonant region, as done conversely in Fig. 9.11a. This allows the definition of the "resonance borders" down to small eccentricity, but with values of semimajor axis which tend respectively to 0 and to infinity as $e \to 0$. Second, they expanded the Hamiltonian in a power series of the eccentricity and retained only the leading term; by doing so, they lost the information on how the collision curve with Jupiter modifies the resonant dynamics, and missed the fact that the resonant region, protected from planetary encounters, shrinks with increasing eccentricity above the planet-crossing limit. Nevertheless, these approximations have only limited importance for practical purposes, so that Dermott and Murray's study remains the key reference on the subject. As can be seen in Fig. 9.12, the mean motion resonances overlap at large eccentricity, roughly

for $Q > 4.2$ AU, where $Q = a(1 + e)$ is the aphelion distance of the asteroid. As seen in Chapter 6, when resonance overlapping occurs, only the central part of the resonances might be expected to host regular motion (the resonant islands), while all the rest of the phase space is globally chaotic. In particular nonresonant regular orbits cannot exist. Only below the threshold for resonant overlapping is the asteroid motion expected to be globally stable, in the framework of the restricted circular planar problem. Chaotic motion can exist, close to the separatrices of the resonances, but the absence of resonance overlapping guarantees the local confinement of the motion, as sketched in Fig. 6.3. Figure 9.12 also plots the osculating semimajor axis and eccentricity of the asteroids known at the time of the study. One notes that the vast majority of the asteroids are located in the regions in between the various mean motion resonances. The mean motion resonant regions are associated with substantial gaps in the asteroid distribution, with the exceptions of the 3/2 resonance and of the 1/1 resonance (the latter located outside of the semimajor axis range covered by the plot). This result confirms the original intuition by Kirkwood (1866), who first noticed the nonuniform distribution of the asteroids, on the basis of the first 91 discoveries. The mean motion resonant regions where asteroids are absent are usually called *Kirkwood gaps.* On the basis of the restricted planar circular problem, the existence of these gaps cannot be understood, because resonant librations are stable. The origin of these gaps will be explained in Chapter 11, when the secular dynamics inside mean motion resonances is studied in the framework of a realistic model of the planetary system. Notice that in Fig. 9.12 a few asteroids seem to be inside the resonances, although close to the resonance borders. Most of these asteroids are actually nonresonant; they appear to be within the separatrices of a resonance because their osculating elements correspond to a value of σ that is different from that used to compute the resonance borders (i.e. $\sigma = \sigma_{\rm stab}$). If their orbital elements were integrated until the condition $\sigma = \sigma_{\rm stab}$ is matched, the resulting values of a and e would put these asteroids outside of the resonant borders in the representation of Fig. 9.12. A consistent computation of the distance of the asteroids from the borders of mean motion resonances will be developed in Section 9.4.

The analytic computation of the widths of mean motion resonances in the Kuiper belt has been done by Morbidelli *et al.* (1995a). Figure 9.13 shows the separatrices of the main resonances with Neptune and with Uranus up to 50 AU. The resonances appear to overlap close to the line $q = a_{\rm N} = 30$ AU, where $q = a(1 - e)$ is the perihelion distance of the small body and $a_{\rm N}$ is the semimajor axis of Neptune. Note that for $q < 30$ AU the widths of the resonant regions shrink with increasing eccentricity, for the reasons explained above in

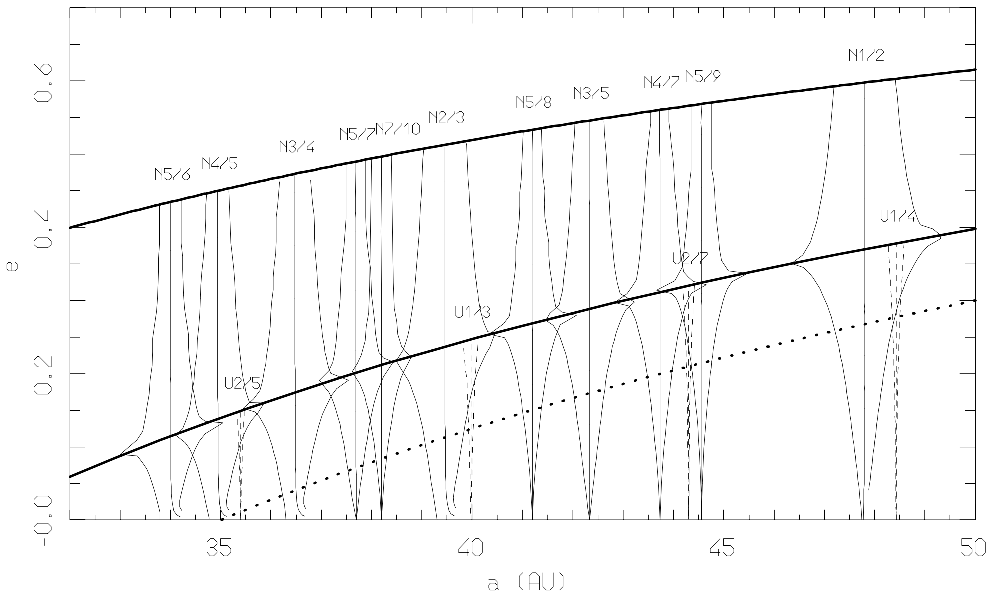


Figure 9.13: The location and the width of mean motion resonances with Neptune (N$n/n_{\rm N}$) and Uranus (U$n/n_{\rm U}$) in the Kuiper belt between 32 and 50 AU. The two bold solid curves correspond to $q = a(1 - e) = 30$ AU and $q = 19.2$ AU, which are respectively the thresholds for Neptune-crossing and Uranus-crossing orbits. The bold dotted curve corresponds to $q = 35$ AU. See text for description. Reprinted from Fig. 1 of Morbidelli *et al.* (1995a), with permission from Academic Press.

this section. These resonant regions are the only regions with $q < 30$ AU that are expected to host regular motion, on the basis of the restricted planar circular problem. Conversely, all nonresonant orbits should be chaotic because they undergo Neptune's encounters. For q somewhat larger than 30 AU, conversely, the bodies do not cross the orbit of Neptune and for $q > 35$ AU the mean motion resonances appear well separated, so that the motion is expected to be globally stable, still in the framework of the planar circular problem. The widths of the regular regions in mean motion resonances with Neptune have also been numerically determined by Malhotra (1996), by computing Poincaré sections of the planar circular restricted problem. The widths reported by Malhotra are slightly smaller than those reported in Fig. 9.13, because they exclude the chaotic regions that exist in correspondence to the separatrices of the resonant normal forms.

An important, but not well quantified, fraction of the bodies in the Kuiper belt librate in the 2/3 resonance with Neptune, similarly to Pluto, and are therefore called *Plutinos*. One other body is believed to be in 3/4 resonance. Conversely, it is not clear yet if the bodies with semimajor axis between 40 and 50 AU are preferentially located within or outside mean motion resonances; the conjectured existence of the equivalent of the Kirkwood gaps in the Kuiper belt has not been established yet.

9.2.1 Threshold for overlapping in the vicinity of the planet

In Fig. 9.13 one can notice that the resonances of first order in the eccentricity (1/2, 2/3, 3/4, 5/6, ...) constitute a sequence that accumulates approaching Neptune; in fact, it is easy to see, using Kepler's law relating the orbital period to the semimajor axis, that the distance between the locations of the resonances of type $k-1/k$ and $k/(k+1)$ is equal to:

$$\Delta a = a_{\jmath}\left|(\frac{k-1}{k})^{\frac{2}{3}} - (\frac{k}{k+1})^{\frac{2}{3}}\right| \sim 2a_{\jmath}/(3k^2) \quad \text{for } k \to +\infty \ , \tag{9.11}$$

where $a_{\jmath}$ is the semimajor axis of the resonant planet. On the other hand, the widths of these resonances do not appear to shrink substantially with the distance to Neptune. This suggests that the resonances of first order in the eccentricity may completely overlap, for k large enough, giving rise to a chaotic region close to the planet that essentially extends down to eccentricity equal to 0. This fact was first shown by Wisdom (1980), who also computed that the threshold on k for the overlap is proportional to $\varepsilon^{-2/7}$, where ε is the mass of the resonant planet. This result holds for both interior and exterior mean motion resonances of first order in the eccentricity.

The derivation of Wisdom's result requires the computation of the width of resonances of type $|k_{\jmath}-k| = 1$, as a function of k. The width depends on the value of N; as shown in Fig. 9.2, if N is small enough the stable equilibrium point is very close to $e = 0$ so that the oscillation of the semimajor axis and the eccentricity along a libration cycle becomes negligible. However, the values of N for which this happens correspond to values of a at $e = 0$ which are very different from the unperturbed location of the resonance $a_{\rm res}$, given by Kepler's law. Therefore, a reasonable choice is to set the value of N such that $a = a_{\rm res}$ corresponding to $e = 0$. From (9.8) this happens for $N = 1/[k^2(k+1)n_{\jmath}]^{1/3}$ for interior resonances and $N = -1/[k^2(k-1)n_{\jmath}]^{1/3}$ for exterior resonances. Let's focus now for simplicity on interior resonances. From (9.6) and (9.7), after expansion in a Taylor series of $\sqrt{S}$, the first terms of the Hamiltonian are:

$$\mathcal{H}_{\rm PC} = -\frac{S}{k^2N^3} - \frac{3S^2}{2k^2N^4} + (k+1)n_{\jmath}S + \varepsilon c\sqrt{S}\cos\sigma \ , \tag{9.12}$$

where all terms independent of S and σ have been dropped, and only the first harmonic has been retained; c is a numerical coefficient, approximately equal to $\sqrt{2}(k+1)/\pi[2K_0(2/3)+K_1(2/3)]$, where K_0 and K_1 are the modified Bessel functions (Murray and Holman, 1997). With the choice of N given above, the linear terms in S annihilate. Then, (9.12) has a unique (stable) equilibrium point at $\sigma_{\rm stab} = 0$, $S_{\rm stab} = (\varepsilon c k^2 N^4/6)^{2/3}$, which is $\sim \varepsilon^{2/3}k^{-2/3}$ (substitute the dependence of c and N on k). The existence of a stable equilibrium point offset from $S = 0$ forces the action S to have oscillation of amplitude $\delta S \sim 2S_{\rm stab}$. The coupled oscillations of the semimajor axis will therefore have amplitude (from the definition of N in formula 9.5) $\delta a = 2\sqrt{a}\delta\Lambda = 2\sqrt{a}k\delta S$. As a consequence, the semimajor axis oscillation due to the resonance, for the considered value of N, is proportional to $\varepsilon^{2/3}k^{1/3}$. Finally, remembering the mutual spacing in the semimajor axis between the resonances of first order in the eccentricity (9.11), one gets that resonance overlapping is possible starting from $k \sim \varepsilon^{-2/7}$.

From Kepler's law, the resonance of first order in the eccentricity with $k \sim \varepsilon^{-2/7}$ occurs for a semimajor axis a such that $|a - a_{\bar{\jmath}}|/a_{\bar{\jmath}} \sim \varepsilon^{2/7}$. This implies that the chaotic region extends on both sides of the planet over a relative semimajor axis interval that is proportional to $\varepsilon^{2/7}$.

9.2.2 Overlapping of resonances with different planets

Mean motion resonances with two different planets can also overlap each other. This is in fact quite common, because quasi-resonant relationships exist among the giant planets. For instance, Jupiter and Saturn are close to the mutual 5/2 mean motion resonance. As a consequence, the n/m mean motion resonances with Jupiter is located close to the $5n/2m$ mean motion resonance with Saturn. A similar case happens for the resonances with Uranus and with Neptune (these two planets are close to the mutual 2/1 resonance). A *Laplace resonance* is a particular case of overlapping mean motion resonances, which occurs when the two perturbers are exactly resonant with each other. Laplace resonances do not occur in asteroid or Kuiper belt dynamics (because the planets are not exactly resonant among themselves), but are quite common in satellite systems. The most famous Laplace resonance occurs for the Galilean satellites, where Europa is simultaneously in the 1/2 resonance with Io and in the 2/1 resonance with Ganymede.

When the two concerned resonances have a comparable width, an effective way to study the consequences of their overlapping is to consider the planar *bicircular problem.* The Hamiltonian of this problem is obtained from (2.38), taking into account only the two resonant planets and assuming them on coplanar and circular orbits. The inclination of the small body is also assumed

to be 0. For the D'Alembert rules, the Hamiltonian depends only on the angles $\lambda, p, \lambda_{j_1}$ and λ_{j_2} (the last two angles being the mean longitudes of the resonant planets), and thus has effectively 4 degrees of freedom. The resonant normal form (see Section 2.5.2) in this case retains from the perturbation only the harmonics that depend solely on the resonant angles, namely $k'\lambda - k_{j_1}\lambda_{j_1} + (k'-k_{j_1})p$ for the k_{j_1}/k' resonance with planet j_1 and $k''\lambda - k_{j_2}\lambda_{j_2} + (k''-k_{j_2})p$ for the k_{j_2}/k'' resonance with planet j_2. All short periodic terms are averaged out in the construction of the normal form.

The first step for the study of the averaged bicircular problem is to find a set of canonical variables, such that two of the new angles are the critical angles of the two resonances. In the case where $k_{j_1} \neq k'$ and $k_{j_2} \neq k''$ the good variables are:

$$\begin{aligned}
\Sigma_1 &= \tfrac{(k_{j_1}-k')(k_{j_2}-k'')}{k''(k_{j_1}-k')-k'(k_{j_2}-k'')}\left(\Lambda + \tfrac{k''}{k_{j_2}-k''}P\right)\ , & \sigma_1 &= \tfrac{k_{j_1}\lambda_{j_1}-k'\lambda}{k_{j_1}-k'} + p \\
\Sigma_2 &= \tfrac{(k_{j_2}-k'')(k_{j_1}-k')}{k'(k_{j_2}-k'')-k''(k_{j_1}-k')}\left(\Lambda + \tfrac{k'}{k_{j_1}-k'}P\right)\ , & \sigma_2 &= \tfrac{k_{j_2}\lambda_{j_2}-k''\lambda}{k_{j_2}-k''} + p \\
\tilde{\Lambda}_{j_1} &= \Lambda_{j_1} - \left(1 + \tfrac{k'}{k_{j_1}-k'}\right)\Sigma_1\ , & \tilde{\lambda}_{j_1} &= \lambda_{j_1} \\
\tilde{\Lambda}_{j_2} &= \Lambda_{j_2} - \left(1 + \tfrac{k''}{k_{j_2}-k''}\right)\Sigma_2\ , & \tilde{\lambda}_{j_2} &= \lambda_{j_2}\ .
\end{aligned} \tag{9.13}$$

If one of the two resonances is a 1/1 resonance (say $k_{j_1} = k'$), the good variables are instead:

$$\begin{aligned}
\Sigma_1 &= \Lambda + \tfrac{k''}{k_{j_2}-k''}P\ , & \sigma_1 &= \lambda - \lambda_{j_1} \\
\Sigma_2 &= P\ , & \sigma_2 &= \tfrac{k_{j_2}\lambda_{j_2}-k''\lambda}{k_{j_2}-k''} + p \\
\tilde{\Lambda}_{j_1} &= \Lambda_{j_1} + \Lambda + \tfrac{k''}{k_{j_2}-k''}P\ , & \tilde{\lambda}_{j_1} &= \lambda_{j_1} \\
\tilde{\Lambda}_{j_2} &= \Lambda_{j_2} - \tfrac{k_{j_2}}{k_{j_2}-k''}P\ , & \tilde{\lambda}_{j_2} &= \lambda_{j_2}\ .
\end{aligned} \tag{9.14}$$

The reader can check that both transformations are canonical, using the Poisson bracket criterion (see Section 1.6). In these new variables, the averaged Hamiltonian of the planar bicircular problem depends only on the angles σ_1 and σ_2. Therefore, it is evident that the problem has effectively only two degrees of freedom. Thus the dynamics can be easily explored using Poincaré sections. Because both angles σ_1 and σ_2 can alternately circulate and librate, it is advisable to compute the section on the angle $k''(k_{j_1}-k')\sigma_1 - k'(k_{j_2}-k'')\sigma_2$ (or $k''\sigma_1 - k'(k_{j_2}-k'')\sigma_2$, if $k_{j_1} = k'$) which typically circulates in a fixed

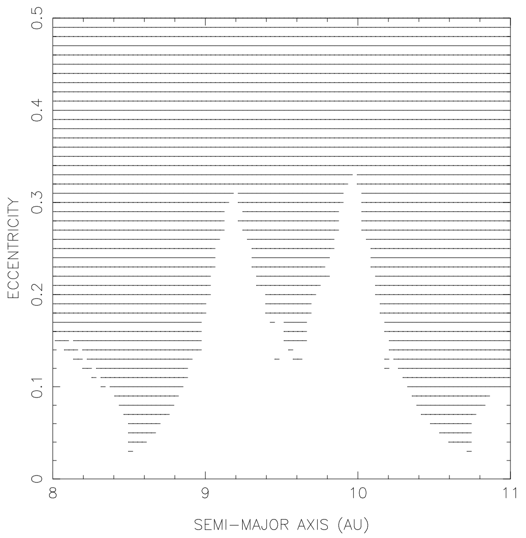


Figure 9.14: The dashed region shows the initial (a, e) that give unstable motion in the averaged bi-circular problem, with harmonics related to the 1/1 resonance with Saturn and the 2/5 resonance with Jupiter. The initial value of the critical angle of the 2/5 resonance with Jupiter is π, while the initial value of the critical angle of the 1/1 resonance with Saturn is the one (dependent on e) that characterizes the center of tadpole librations in the integrable approximation where Jupiter is not taken into account. Courtesy of D. Nesvorný.

direction.[3]

Perhaps, the most interesting application of this approach is to the study of the overlap between the 1/1 resonance with Saturn and the 2/5 resonance with Jupiter. The two resonances have comparable width, if the eccentricity of the asteroid is about 0.1. De la Barre *et al.* (1996) claimed that this overlap is responsible for the absence of Saturn's Trojans. The shaded region in Fig. 9.14 represents the set of initial conditions in the (a, e) plane that lead to unstable motion in the integration of the averaged bicircular problem. By unstable motion we mean here motion that escapes from the domain of definition of the Hamiltonian, namely that crosses the collision line with one of the planets.

[3]In fact, this combination of σ_1 and σ_2 is equal to $k''k_{j_1}\lambda_{j_1} - k'k_{j_2}\lambda_{j_2} + [k''(k_{j_1} - k') - k'(k_{j_2} - k'')]p$. Because λ_{j_1} and λ_{j_2} are not exactly resonant their linear combination circulates with time. On the other hand, the motion of p is typically too slow to change the direction of the motion. This is of course not true for a Laplace resonance.

As one sees, the orbits at the center of Saturn's tadpoles region ($a \sim 9.5$ AU) with $e > 0.12$ are unstable. These orbits would be stable if only Saturn were taken into account; therefore their instability *must* be due to the overlap with the 2/5 resonance with Jupiter. The orbits with the same semimajor axis but with initial eccentricity smaller than 0.12 are stable in the bicircular problem. However, we will see in Section 11.5 that the secular dynamics drives their eccentricity above 0.12, forcing them to enter the instability region.

As anticipated above, the study of the averaged bicircular problem is suitable in the case where the two resonances have comparable width, so that their overlap leads to strong chaos. But when one resonance is much thinner than the other, as in the case of the thin 5/1 resonance with Saturn and of the large 2/1 resonance with Jupiter, the effects of the overlap are very subtle, and the bicircular model may be a too rough an approximation of the real dynamics. In these cases it is more effective to study the dynamics with a hierarchical perturbation approach, as explained in Section 11.2.5.

9.3 Resonant multiplets

We now come back to the dynamics of an isolated mean motion resonance, but we leave the integrable approximation (9.4) studied so far and reintroduce the terms dropped from (9.1), namely the inclination of the asteroid, and the eccentricities and inclinations of the planets. For completeness, we also take into account the secular precession of planetary orbits; for this purpose, as in Chapter 8, assuming that the planetary elements evolve according to the Lagrange–Laplace solution (7.10), we introduce new angles $\varpi_j^* = g_j t + \beta_j$ and $\Omega_j^* = s_j t + \delta_j$, and denote by Λ_{g_j} and Λ_{s_j} the conjugate actions ($j = 1, \ldots, n$). To make the Hamiltonian autonomous, the term $\sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j})$ is added to the integrable part of the Hamiltonian; remember that the planetary precession frequencies are slow, so that g_j and s_j are no larger than ε. With these settings the mean motion resonant normal form (9.1) can be rewritten as:

$$\mathcal{H}_{\mathrm{MMR}} = \mathcal{H}_0(\Lambda, \Lambda_{\bar{j}}, \Lambda_{g_j}, \Lambda_{s_j}) + \varepsilon \mathcal{H}_1(\Lambda, P, Q, p, q, k_{\bar{j}}\lambda_{\bar{j}} - k\lambda, \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*) , \quad (9.15)$$

where $\boldsymbol{\varpi}^*$ and $\boldsymbol{\Omega}^*$ denote vectors with components ϖ_j^* and Ω_j^* respectively. In the expression above one has

$$\mathcal{H}_0 = -\frac{1}{2\Lambda^2} + n_{\bar{j}}\Lambda_{\bar{j}} + \sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j}) , \quad (9.16)$$

while the perturbation $\varepsilon\mathcal{H}_1$ has Fourier expansion:

$$\varepsilon\mathcal{H}_1 = \sum_{m,\mathbf{u},\mathbf{v},s,r} c_{m,\mathbf{u},\mathbf{v},s,r}(\Lambda, P, Q) \exp\{\iota[m(k_{\bar{j}}\lambda_{\bar{j}}-k\lambda)+\mathbf{u}\cdot\boldsymbol{\varpi}^*+\mathbf{v}\cdot\boldsymbol{\Omega}^*+sp+rq]\} \ , \quad (9.17)$$

where m, r, s are integer numbers and $\mathbf{u} \equiv (u_1, \ldots, u_n)$ and $\mathbf{v} \equiv (v_1, \ldots, v_n)$ are vectors with integer components, such that $m(k_{\bar{j}}-k)-s-r+\sum_j(u_j+v_j) = 0$. The coefficients $c_{m,\mathbf{u},\mathbf{v},s,r}$ depend parametrically on the amplitudes $M_{j,k}, N_{j,k}$ of the planetary secular motion that appear in (7.10).

In the expansion (9.17) we distinguish the harmonics with $m = 0$ from those with $m \neq 0$. The first ones are independent of the fast angles λ and $\lambda_{\bar{j}}$ and were already present in the problem of the asteroid's secular motion, studied in Chapter 8. They rule over the secular evolution of the asteroid's eccentricity and inclination. The terms with $m \neq 0$ are, properly speaking, the mean motion resonant harmonics, because they all contain the resonant combination $k_{\bar{j}}\lambda_{\bar{j}} - k\lambda$.

Let us neglect for the moment the secular harmonics ($m = 0$) and analyze the structure of the dynamical system that results from the presence of the sole mean motion resonant harmonics. The system is highly nonintegrable, because it depends on several independent and quasi-resonant combinations of the angles.

To study this multiresonant dynamics, we first introduce the angles $\sigma_{m,\mathbf{u},\mathbf{v},s,r} = m(k_{\bar{j}}\lambda_{\bar{j}} - k\lambda) + \mathbf{u}\cdot\boldsymbol{\varpi}^* + \mathbf{v}\cdot\boldsymbol{\Omega}^* + sp + rq$. Each mean motion resonant harmonic has for its argument a different angle $\sigma_{m,\mathbf{u},\mathbf{v},s,r}$. Then, we consider the Hamiltonians

$$\mathcal{H}_{m,\mathbf{u},\mathbf{v},s,r} = \mathcal{H}_0(\Lambda, \Lambda_{\bar{j}}, \Lambda_{g_j}, \Lambda_{s_j}) + c_{\mathbf{0}} + c_{m,\mathbf{u},\mathbf{v},s,r}\cos\sigma_{m,\mathbf{u},\mathbf{v},s,r} \quad (9.18)$$

where, for simplicity of notation, we have denoted by $c_{\mathbf{0}}$ the coefficient $c_{0,\mathbf{0},\mathbf{0},0,0}$ and we have emphasized the fact that the Fourier expansion (9.17) contains only cosine terms. These Hamiltonians play the role of the single-resonance models used in Chapter 6 to study the interaction among resonances. For each Hamiltonian $\mathcal{H}_{m,\mathbf{u},\mathbf{v},s,r}$ the exact resonance occurs for $\dot{\sigma}_{m,\mathbf{u},\mathbf{v},s,r} = 0$, i.e. for

$$m\left[k_{\bar{j}}n_{\bar{j}} - k\left(\frac{1}{\Lambda^3} + \frac{\partial c_{\mathbf{0}}}{\partial\Lambda}\right)\right] + s\frac{\partial c_{\mathbf{0}}}{\partial P} + r\frac{\partial c_{\mathbf{0}}}{\partial Q} + \sum_j(u_j g_j + v_j s_j) = 0 \ . \quad (9.19)$$

This equation is solved for a value of Λ that depends on the actions P and Q and also on the indexes $m, \mathbf{u}, \mathbf{v}, s, r$; however, because $c_{\mathbf{0}}, g_j$ and s_j are all of order ε, the solution must be ε-close to $\Lambda = (k_{\bar{j}}n_{\bar{j}}/k)^{1/3}$, i.e. to the unperturbed location of the mean motion resonance given by Kepler's law. Therefore, we conclude that the mean motion resonance corresponds to a

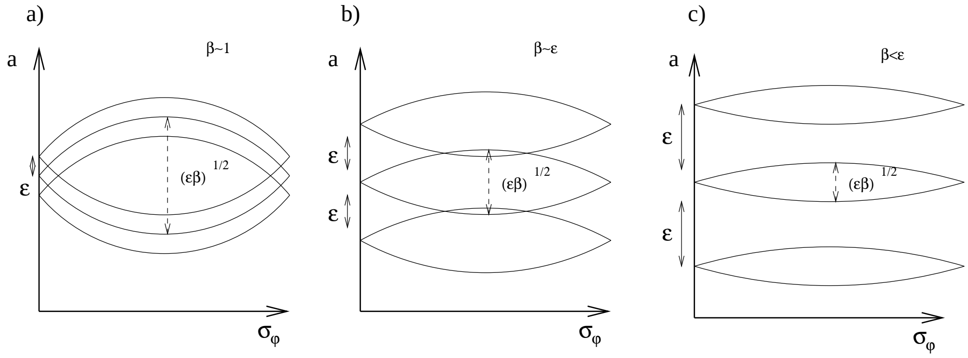


Figure 9.15: Sketch of the possible dynamical structures inside a mean motion resonant multiplet. (a): complete overlap, leading to modulated pendulum-like dynamics; (b): marginal overlap; (c): separation of the resonances in the multiplet. The label σ_φ generically denotes the angles $\sigma_{m,\mathbf{u},\mathbf{v},s,r}$ of the various resonances in the multiplet. Reprinted from Fig. 13 of Morbidelli and Guzzo (1996), with permission from Kluwer Academic Publishers.

multiplet of resonances whose distinct locations are ε-close in the asteroid's semimajor axis. This multiplet structure is a specific property of mean motion resonances and, more generally, of the resonances among the fast angles in a degenerate system (Guzzo and Morbidelli, 1997).

To compute the width in the semimajor axis of each resonance in the multiplet we refer again to the corresponding integrable model (9.18); the width is computed, like in Chapters 4 and 6, as the maximal distance between the separatrices, and it is of order $\sqrt{c_{m,\mathbf{u},\mathbf{v},s,r}}$. Remember now that the coefficients $c_{m,\mathbf{u},\mathbf{v},s,r}$ cannot be larger than order ε (they are the coefficients of the Fourier expansion of $\varepsilon\mathcal{H}_1$), but they can possibly be much smaller. Denote by $\varepsilon\beta$ the size of the largest coefficient(s). Then, the possibilities are those sketched in Fig. 9.15: if $\beta \sim 1$, the major resonances in the multiplet have widths that are much larger than their mutual separation (respectively $\sim \sqrt{\varepsilon}$ and $\sim \varepsilon$) so that there is a situation of *complete* resonance overlap (panel a); if $\beta \sim \varepsilon$ the major resonances have widths of the same order as their mutual separation ($\sim \varepsilon$), so that *partial overlap* occurs (panel b); finally, if $\beta << \varepsilon$, all the resonances are much thinner than their mutual separation, so that we expect no overlap among the resonances in the multiplet (panel c).

In mean motion resonances with Jupiter, the values of $|k - k_{\jmath}|$ and of the asteroid's eccentricity and inclination dictate which of the three situations sketched above must occur. In fact, for the D'Alembert rules, each coefficient

$c_{m,\mathbf{u},\mathbf{v},s,r}$ is equal to $\varepsilon e^{|s|} i^{|r|} \prod_j M_{5,j}^{|u_j|} N_{5,j}^{|v_j|}$ times some numerical factor (not necessarily of order unity), where $M_{5,j}$ and $N_{5,j}$ are the constants entering in (7.10) for the secular motion of Jupiter's elements. Moreover, $s + r - \sum_j (u_j + v_j) = m(k - k_{\bar{j}})$. As a consequence, for mean motion resonances of low eccentricity order (i.e. $|k - k_{\bar{j}}|$ of order unity), some of the coefficients $c_{m,\mathbf{u},\mathbf{v},s,r}$ must be proportional to powers of e, i, $M_{5,j}$ or $N_{5,j}$ that are of order unity, and therefore cannot be significantly smaller than ε (remember from formula 8.7 that e and i oscillate due to the planetary secular perturbations so that they cannot be assumed significantly smaller than Jupiter's eccentricity and inclination). Therefore, mean motion resonances with Jupiter of low order in the eccentricity are expected to correspond to the situation of Fig. 9.15a. This dynamical structure, equivalent to that of a *modulated pendulum*, will be discussed in the next section. The situation changes gradually with increasing values of $|k - k_{\bar{j}}|$: the coefficients $c_{m,\mathbf{u},\mathbf{v},s,r}$ become proportional to increasing powers of the eccentricities and inclinations, so that the situations sketched in Fig. 9.15b and c are attained, first for values of e and i of the same order of Jupiter's eccentricity and inclination, and then also for larger values.

However, for mean motion resonances with planets of much smaller size than Jupiter, the situation of marginal overlap among the resonances in the multiplet can be attained for smaller $|k - k_{\bar{j}}|$. In fact, from (9.19) the separation among the resonances in the multiplet is of order ε (the mass of Jupiter relative to that of the Sun) whatever the resonant planet; conversely the coefficients $c_{m,\mathbf{u},\mathbf{v},s,r}$ are, for these resonances, proportional to the mass of the resonant planet $\varepsilon_{\bar{j}} \ll \varepsilon$.

The situation of Fig. 9.15b is already expected for very low eccentricity orders in the case of three-body resonances, namely for the resonances among the mean motions of an asteroid and of two planets. In fact, for these resonances, the coefficients $c_{m,\mathbf{u},\mathbf{v},s,r}$ are of order ε^2, because the resonant harmonics appear in the normal form only at order 2, as explained in Section 2.5.2. Three-body mean motion resonances will be discussed in detail in Chapter 10.

The precise understanding of the multiplet structure requires of course the quantitative computation of the separatrices of each single-resonance Hamiltonian (9.18). Nevertheless, a simple diagnostic can be done on the basis of a short numerical integration. Because the separation of the resonances in the multiplet is proportional to the secular precession frequencies (see equation 9.19), while the amplitude of a resonance is proportional to the libration frequency of its critical argument $\sigma_{m,\mathbf{u},\mathbf{v},s,r}$ (see Chapter 4), the case of Fig. 9.15a should be expected when the frequency of libration is much faster than the secular precession frequencies. Conversely, the case of Fig. 9.15b requires that the frequencies are of the same order, while the case of Fig. 9.15c

can occur only when the frequencies of libration of all angles $\sigma_{m,\mathbf{u},\mathbf{v},s,r}$ are much slower than the secular frequencies.

9.4 The modulated pendulum approximation

The time-dependent one-degree of freedom Hamiltonian of the modulated pendulum is

$$\mathcal{H} = \frac{\beta}{2}p^2 + c(1 + \alpha\cos(\varepsilon t))\cos q \ , \tag{9.20}$$

where p, q are generic action–angle variables, t is the time, ε is a small parameter and α, β and c are numerical coefficients. The modulated pendulum is the paradigm model for a system of resonances that completely overlap, as in Fig. 9.15a. In fact, with some trigonometrical transformations, it is easy to see that (9.20) is equivalent to the two-degrees of freedom Hamiltonian

$$\mathcal{H} = \frac{\beta}{2}p^2 + \varepsilon\mathcal{T} + c\left(\cos q + \frac{\alpha}{2}(\cos(q-\tau) + \cos(q+\tau))\right) \ , \tag{9.21}$$

where $\mathcal{T}, \tau \equiv \varepsilon t$ are conjugate action–angle variables. The Hamiltonian (9.21) presents a multiplet of three resonances, with critical angles q, $q - \tau$ and $q + \tau$, which are ε-close (they are located respectively at $p = 0$ and $p = \pm\varepsilon$) and whose half-widths are large with respect to ε (respectively $2\sqrt{c/\beta}$ and $2\sqrt{c\alpha/\beta}$).

The Hamiltonian (9.20) has been extensively studied both analytically (see for instance Elskens and Escande, 1991; Bruhwiler and Cary, 1989) and numerically (e.g. Henrard and Henrard, 1991). Figure 9.16 shows its Poincaré section. The pictures are similar to the phase portrait of a pendulum, but instead of a separatrix they show the existence of a large chaotic layer, where regular islands are small with ε. As one sees, in panel (a) the *outer* boundary of the chaotic layer is a curve which is basically equal[4] to the separatrix of a pendulum with Hamiltonian $\beta p^2/2 + c(1 + \alpha)\cos q$; analogously, in panel (b), the *inner* boundary of the chaotic layer is a curve that almost coincides with the separatrix of a pendulum with Hamiltonian $\beta p^2/2 + c(1 - \alpha)\cos q$. This means that the chaotic layer covers the region that is spanned by the instantaneous separatrices of (9.20) during a period of modulation of $c(1 + \alpha\cos(\varepsilon t))$, where by *instantaneous separatrices* we intend those computed from (9.20) assuming t as a constant. As a consequence, a regular region filled by librational invariant tori exists if and only if $\alpha < 1$.

Of course, the dynamics in a real mean motion resonant multiplet, even in the case of complete overlap, is more complicated than the dynamics of

[4]The equality holds in the limit $\varepsilon \to 0$.

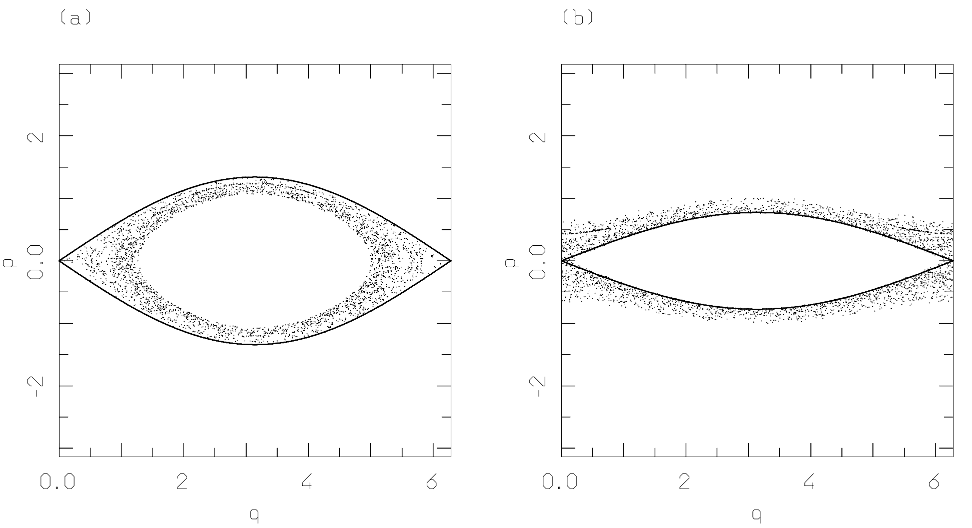


Figure 9.16: Poincare sections of (9.20) for $\varepsilon t = 0 \bmod 2\pi$ (panel a) and for $\varepsilon t = \pi \bmod 2\pi$ (panel b). The solid curve denotes the separatrix of the pendulum obtained from (9.20) under the assumption that εt is a fixed parameter equal to 0 (a) or π (b).

(9.20). In fact, the mean motion resonant multiplet has in general more than three components and, moreover, it does not have the nice symmetry of the multiplet of (9.21). Nevertheless one can still compute with good accuracy the extent of the chaotic layer using a recipe equivalent to that used for the modulated pendulum. Using the variables (9.5), one rewrites (9.15) as

$$\mathcal{H}_{\rm MMR} = \mathcal{H}_0(N, S, S_z, \Lambda_{g_j}, \Lambda_{s_j}) + \varepsilon\mathcal{H}_1(S, N, S_z, \sigma, \nu, \sigma_z, \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*) \ , \qquad (9.22)$$

and calculates its instantaneous separatrices under the assumption that the secular angles $\nu, \sigma_z, \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*$ and the actions N, S_z are constant parameters (so that $\mathcal{H}_{\rm MMR}$ depends only on S, σ and is thus integrable): the chaotic layer should cover the region spanned by the instantaneous separatrices for changing values of the secular angles, and its boundaries should be given by the instantaneous separatrices respectively corresponding to the minimal and to the maximal resonant width. As an example, Fig. 9.17 shows these extreme instantaneous separatrices for the case of the 5/2 mean motion resonance with Jupiter, in a simplified model where the asteroid and the planet are assumed to be in coplanar orbits and the orbit of Jupiter is elliptic and fixed. Analogous figures for different values of the secular angles and for various mean motion resonances can be found in Yoshikawa (1990, 1991). As one sees in Fig. 9.17,

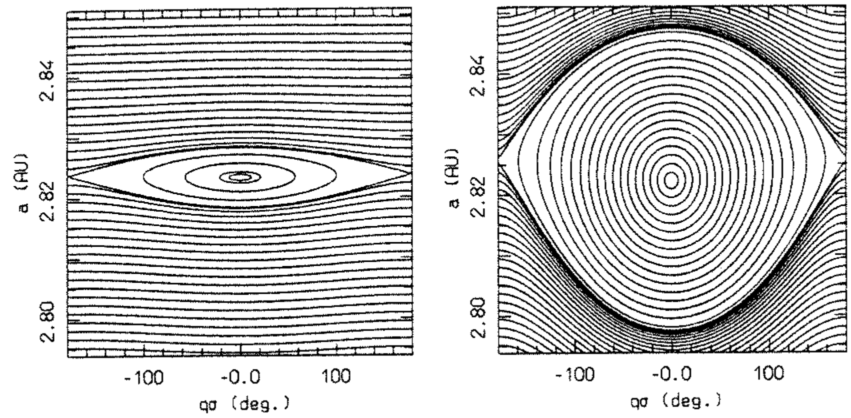


Figure 9.17: Dynamics in the 5/2 mean motion resonance with Jupiter, computed for $e_{\text{J}} = 0.05989$, a value of N that corresponds to $e = 0.204$ at $a = 2.825$ AU, and two different values of $\varpi - \varpi_{\text{J}}$: 0 (left panel) and π (right panel). The label $q\sigma$ on the x-axis stands for $(k_{\bar{J}} - k)\sigma$, which is equal to 3σ for the 5/2 resonance. Reprinted from Fig. 2 of Morbidelli *et al.* (1995b), with permission from Academic Press.

the width of the region of libration is much smaller in the minimal case (left panel) than in the maximal case (right panel), but it is never zero; therefore a regular libration region is expected, surrounded by a chaotic layer, like for the modulated pendulum in the case $\alpha < 1$. In general, a regular libration region is expected whenever the asteroid's inclination is moderate and its eccentricity is significantly larger than the planetary eccentricities, because in these cases – for the D'Alembert rules – the width of the resonance related to the planar circular restricted problem (9.4) dominates over the widths of all other resonances in the multiplet. The adjective "regular" should not be misinterpreted here: the region is regular only in the oversimplified model where the asteroid's actions N, S_z are assumed to be constant. In reality these actions evolve in a secular motion that may still be unstable and chaotic. The study of the secular dynamics in these "regular" regions is one of the most complicated issues in Celestial Mechanics, as will be discussed in Chapter 11.

The computation of the boundaries of the chaotic layer of a mean motion resonance is not enough to understand whether a given asteroid is outside of the resonance, in the chaotic layer, or in the "regular" libration region. The reason is that the precession of the secular angles, which causes the modulation of the mean motion resonance width, also forces the oscillation of the asteroid's

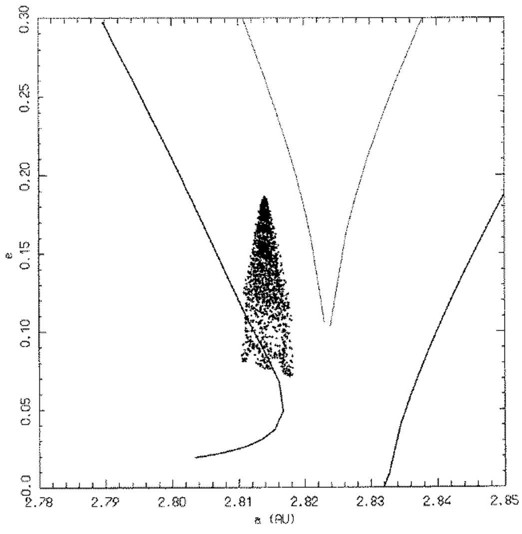


Figure 9.18: Numerical integration of the equations of the normal form for the 5/2 mean motion resonance with Jupiter. The secular variations of Jupiter's orbit with frequencies g_5 and g_6 are taken into account. See text for comments. Reprinted from Fig. 3 of Morbidelli *et al.* (1995b), with permission from Academic Press.

eccentricity and inclination; this in turn changes the width of the resonance, since the latter depends also on the values of both e and i. Figure 9.18 shows an example of these competing effects. The model now accounts for the terms with frequencies g_5 and g_6 in the equations (7.10) for the secular motion of Jupiter's eccentricity and perihelion longitude. The two light curves denote the minimal width in the (a, e) plane of the 5/2 resonance, which corresponds to $\varpi - \varpi_5^* = 0$ and $\varpi - \varpi_6^* = \pi$; the two bold curves denote the maximal width of the resonance, related to $\varpi - \varpi_5^* = \pi$ and $\varpi - \varpi_6^* = 0$. The dots trace the evolution of an asteroid obtained by numerically integrating the equations of motion of the resonant normal form (9.15). During this integration, the asteroid slowly goes up and down on the (a, e) plot because of the secular evolution of its eccentricity. In the meantime, it performs rapid and almost horizontal oscillations that are due to the variations of the semimajor axis, forced by the proximity of the resonance and correlated with the circulation of σ. The maximal eccentricity is attained when $\varpi - \varpi_5^* = 0$ and $\varpi - \varpi_6^* = \pi$, and at that instant the position of the asteroid should be compared with the light borders of the resonance; the minimal eccentricity is achieved when $\varpi - \varpi_5^* = \pi$ and $\varpi - \varpi_6^* = 0$, namely when the resonance width is maximal. The cone-like

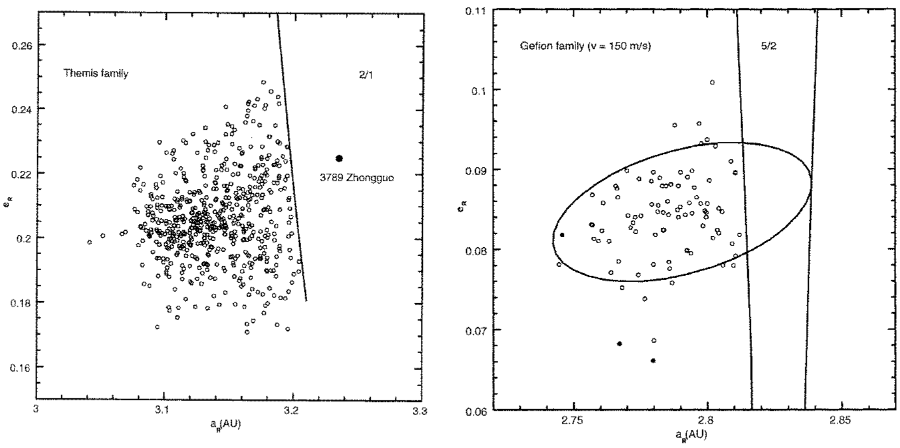


Figure 9.19: Left panel: the position of the asteroids of the Themis family and the border of the chaotic region of the 2/1 resonance with Jupiter; the position of the asteroid Zhonnguo, one of the few long-living bodies in the 2/1 resonance (see Chapter 11) is also indicated. Right panel: the same, but for the Gefion family and the 5/2 resonance; the ellipse indicates the expected distribution of the Gefion family members according to a collisional model. The definition of the elements a_R, e_R that allow a precise comparison of the asteroids' positions vs. the borders of the resonance is explained in the text. Reprinted from Figs. 6 and 17 of Morbidelli *et al.* (1995b), with permission from Academic Press.

shape left by the trace of the asteroid's evolution indicates that the maximal amplitude of oscillation of the semimajor axis occurs when the eccentricity is at the minimum of its secular cycle; but from Fig. 9.17 we know that the more a oscillates, the closer is the asteroid to the resonance's separatrix. Therefore, we can conclude that the closest approach of the asteroid to the modulated resonance happens when $\varpi - \varpi_5^* = \pi$ and $\varpi - \varpi_6^* = 0$. The fact that the bottom left corner of the cone-like shape is outside of the region bounded by the bold curves implies that the asteroid is outside of the resonance also during the phase of closest approach; in the opposite case the trace left by the oscillation of the semimajor axis would lie within the bold curves and would be centered around ~ 2.825 AU. Thus we can conclude that this asteroid is outside of the chaotic layer of the 5/2 mean motion resonance.

If we want to unambiguously represent the position of this asteroid relative to the 5/2 resonance, the natural choice is to use for the asteroid the values of the semimajor axis and eccentricity, say (a_R, e_R), that correspond to the bot-

tom left corner of the cone-like shape of Fig. 9.18, and to plot the boundaries of the chaotic layer given by the bold separatrices. In this way we represent the asteroid and the separatrices of the resonance using the same values of all the angles $(\sigma, \varpi, \varpi_5^*, \varpi_6^*)$, and at the instant of their closest mutual approach.

However, the dynamics is not like that illustrated in Fig. 9.18 for all the mean motion resonances, because it depends on the *local* amplitude of the secular oscillation of the asteroid's eccentricity and on the amplitude of modulation of the resonance. For instance, for the 2/1 resonance with Jupiter the closest approach between the asteroid and the resonance's chaotic layer occurs when $\varpi - \varpi_5^* = 0$ and $\varpi - \varpi_6^* = \pi$, namely when the asteroid is at the top of its secular eccentricity cycle and the amplitude of the resonance is minimal. For the 3/1 resonance, the closest approach happens in an intermediate situation, when $\varpi - \varpi_5^* = \varpi - \varpi_6^* = \pi$. Therefore, each resonance requires different phases of the secular angles for a correct representation of the position of the asteroids relative to its boundaries.

Morbidelli *et al.* (1995b) have done the first systematic study of the position of the asteroids of the major asteroid families with respect to the boundaries of the chaotic layers of the nearby resonances. Figure 9.19 shows the results for the Themis and the Gefion families, respectively close to the 2/1 and 5/2 resonances with Jupiter. As one notices, all asteroids are outside the border of the resonant chaotic layer. Actually, the distributions of the asteroids of these families appear truncated at the exact position of the resonance border. This suggests that, at the time of the formation of these families by the catastrophic break-up of the parent bodies, several asteroids were injected into resonance. These bodies must have been subsequently eliminated by the dynamical action of the resonances. Because the asteroid families should be much younger than the age of the Solar System (as can be deduced from collisional studies; Marzari *et al.*, 1995, 1999), this indicates that the formation of the Kirkwood gaps cannot be due to some phenomenon that occurred during the very early phase of the Solar System, but it must be due to "some" dynamical instability mechanism, still ongoing in the present-day system (see Chapter 11).

Chapter 10

THREE-BODY RESONANCES

10.1 Origin of the resonant perturbation terms

The three-body resonances belong to the general class of mean motion resonances, but, differently from the *two body* mean motion resonances discussed in Chapter 11, they involve the orbital frequencies of three bodies. More specifically, a three-body resonance corresponds to the relationship

$$k_1\dot{\lambda}_1 + k_2\dot{\lambda}_2 + k_3\dot{\lambda}_3 \sim 0 \ , \tag{10.1}$$

where $\dot{\lambda}_1, \dot{\lambda}_2, \dot{\lambda}_3$ denote the mean motions of three different bodies and k_1, k_2, k_3 are nonzero integers. In this chapter, we require that (10.1) cannot be decomposed into

$$k_1\dot{\lambda}_1 + (k_2 - m)\dot{\lambda}_2 \sim 0 \ ; \quad m\dot{\lambda}_2 + k_3\dot{\lambda}_3 \sim 0 \ , \tag{10.2}$$

for some integer m. In the opposite case, the three-body resonance is just a consequence of the overlapping of two independent two-body mean motion resonances (called *Laplace resonance* in the case of exact overlapping) and can be studied as explained in Sections 9.2.2 and 11.2.5.

For their importance in planetary dynamics, as well as in the dynamics of the asteroid and of the Kuiper belts, the three-body resonances merit a chapter by themselves. As far as the properties of motion are concerned, three-body resonances do not have specific differences with respect to two-body resonances but, the coefficients of the resonant harmonics being quadratic in the planetary masses, they better show the *multiplet structure* discussed in the previous chapter. Moreover, it is less trivial to understand the origin of the

harmonics associated to three-body resonances and compute their coefficients, because the original Hamiltonian (see 1.29 for the asteroid problem and 1.38 for the planetary problem), once written in modified Delaunay variables (see respectively 2.38 and 2.39), does not present harmonics depending on the combination of the mean longitudes of three bodies. Therefore, one could naively think that three-body resonances have no effect because the corresponding resonant harmonics have null coefficients. Of course, this is not true. In this section, we discuss how the resonant harmonics are generated. We distinguish between a *direct effect*, common to both the restricted problem and the planetary problem, and an *indirect effect*, specific to the restricted case.

10.1.1 The direct effect

The harmonics depending on the combination of the mean longitudes of three bodies are just hidden in the original Hamiltonian (2.38) or (2.39), and they appear at the second order in the planetary masses when the Hamiltonian is averaged over the short periodic terms using a perturbation algorithm. This has already been anticipated in Section 2.5, and we detail this process here. The restricted problem and the planetary problem are formally equivalent, and we focus on the former for simplicity of notation.

We start from the Hamiltonian (2.38), namely

$$\mathcal{H} = \mathcal{H}_0 + \varepsilon\mathcal{H}_1 = -\frac{1}{2\Lambda^2} + \sum_{j=1}^{N} n_j\Lambda_j + \varepsilon\mathcal{H}_1 \ , \tag{10.3}$$

with

$$\varepsilon\mathcal{H}_1 = \sum_{j=1}^{N} \varepsilon_j \mathcal{H}_1^{(j)}(\Lambda, P, Q, \lambda, p, q, \lambda_j, p_j, q_j) \ , \tag{10.4}$$

where n_j is the orbital frequency of the j-th planet of mass ε_j and N is the number of considered planets. Recall that all planets are considered to move on fixed Keplerian ellipses, so that p_j, q_j can be considered as parameters,[1] while the actions Λ_j, conjugate to λ_j, are introduced in order to make the Hamiltonian time independent.

Under the assumption that the mean motion of the asteroid in not in two-body resonance with any of the mean motions of the planets, all harmonics depending on λ, λ_j in (10.4) can be averaged out at order ε. As explained in Section 2.5, the averaging of the perturbation $\mathcal{H}_1$ over the fast angles λ, λ_j, is done following Lie's algorithm, by introducing a generating Hamiltonian χ

[1]Taking into account the slow precession of p_j, q_j, would not change what follows.

such that

$$\{\mathcal{H}_0,\chi\}+\mathcal{H}_1=\overline{\mathcal{H}}_1\equiv\frac{1}{(2\pi)^2}\sum_{j=1}^{N}\int_0^{2\pi}\int_0^{2\pi}\mathcal{H}_1^{(j)}\mathrm{d}\lambda\mathrm{d}\lambda_j\ , \tag{10.5}$$

where $\{f,g\}$ denotes the Poisson bracket of functions f and g. The solution of equation (10.5) has the form:

$$\varepsilon\chi=\sum_{j=1}^{N}\varepsilon_j\chi^{(j)}\ , \tag{10.6}$$

where each $\chi^{(j)}$ is a function only of the variables related to the asteroid and to the j-th planet.

With the canonical variables $(\Lambda^1,P^1,Q^1,\Lambda^1_j,\lambda^1,p^1,q^1,\lambda^1_j)$ (the semimean modified Delaunay variables) introduced by the Lie series S^ε_χ, namely $\Lambda=S^\varepsilon_\chi\Lambda^1,\ldots,\lambda=S^\varepsilon_\chi\lambda^1,\ldots$, the new Hamiltonian becomes:

$$\mathcal{H}^1=\mathcal{H}_0+\varepsilon\left(\mathcal{H}_1+\{\mathcal{H}_0,\chi\}\right)+\varepsilon^2\left(\frac{1}{2}\{\{\mathcal{H}_0,\chi\},\chi\}+\{\mathcal{H}_1,\chi\}\right)+O(\varepsilon^3)\ . \tag{10.7}$$

Since, by construction, $\{\mathcal{H}_0,\chi\}=\overline{\mathcal{H}}_1-\mathcal{H}_1$, the Hamiltonian (10.7) can be rewritten as

$$\mathcal{H}^1=\mathcal{H}_0+\varepsilon\overline{\mathcal{H}}_1+\frac{\varepsilon^2}{2}\left(\{\mathcal{H}_1,\chi\}+\{\overline{\mathcal{H}}_1,\chi\}\right)+O(\varepsilon^3)\ . \tag{10.8}$$

Recalling the expressions of $\mathcal{H}_1$, $\overline{\mathcal{H}}_1$ and χ in (10.4) (10.5) and (10.6) and the definition of Poisson bracket, it is immediately evident that $\varepsilon^2\{\mathcal{H}_1,\chi\}$ and $\varepsilon^2\{\overline{\mathcal{H}}_1,\chi\}$ have the form:

$$\sum_{j=1}^{N}\sum_{k=1}^{N}\Big[\,\varepsilon_j\varepsilon_k\frac{\partial\mathcal{F}_j}{\partial\varphi}(I,\varphi,\varphi_j)\cdot\frac{\partial\chi^{(k)}}{\partial I}(I,\varphi,\varphi_k) \\ -\varepsilon_j\varepsilon_k\frac{\partial\mathcal{F}_j}{\partial I}(I,\varphi,\varphi_j)\cdot\frac{\partial\chi^{(k)}}{\partial\varphi}(I,\varphi,\varphi_k)\Big]\ , \tag{10.9}$$

where $\mathcal{F}_j$ stands for $\mathcal{H}_1^{(j)}$ or $\overline{\mathcal{H}}_1^{(j)}$, the variables (I,φ) generically denote the canonical action–angle variables of the asteroid, (φ_j,φ_k) denote the angles of the j-th and of the k-th planet respectively, and the expression $\frac{\partial}{\partial I}\cdot\frac{\partial}{\partial\varphi}$ denotes the scalar product between the two gradients. Therefore $\varepsilon^2\{\mathcal{H}_1,\chi\}$ and $\varepsilon^2\{\overline{\mathcal{H}}_1,\chi\}$ contain terms which couple the variables of the j-th and of the k-th planet. The harmonics which depend on a combination of the mean

longitudes of the asteroid and of two planets, and which may be related to a three-body resonance, appear in $\varepsilon^2\{\mathcal{H}_1, \chi\}$. Their coefficients are proportional to the product of the masses of the resonant planets $\varepsilon_j \varepsilon_k$.

To construct a three-body resonant normal form to second order in the planetary masses, the Hamiltonian (10.8) is then averaged over all nonresonant combinations of the fast angles λ^1, λ_j^1 ($j = 1, \ldots, N$), retaining only the secular harmonics and the resonant harmonics. In principle, this requires the introduction of a new generating Hamiltonian χ_2 and of new variables $(\Lambda^2, P^2, Q^2, \Lambda_j^2, \lambda^2, p^2, q^2, \lambda_j^2)$ through the canonical transformation $\Lambda^1 = S_{\chi_2}^{\varepsilon^2} \Lambda^2, \ldots, \lambda^1 = S_{\chi_2}^{\varepsilon^2} \lambda^2, \ldots$. As usual (see Section 2.5.2), the generating Hamiltonian is chosen such that

$$\{\mathcal{H}_0, \chi_2\} + \mathcal{H}_2 = \bar{\mathcal{H}}_2 \tag{10.10}$$

where $\mathcal{H}_2 = \frac{1}{2}(\{\mathcal{H}_1, \chi\} + \{\overline{\mathcal{H}}_1, \chi\})$ and $\bar{\mathcal{H}}_2$ is a function which retains only the harmonics of $\mathcal{H}_2$ of type

$$\exp\left[\iota(k_1\lambda_{j_1}^2 + k_2\lambda_{j_2}^2 + k\lambda^2 + m_1 p_{j_1} + m_2 p_{j_2} + r_1 q_{j_1} + r_2 q_{j_2} + mp^2 + rq^2\right] \tag{10.11}$$

with integers k_1, k_2, k equal to zero (the secular terms in the Hamiltonian) or satisfying the resonant relationship $k_1 n_{j_1} + k_2 n_{j_2} + k/\Lambda^3 \sim 0$, where n_{j_1} and n_{j_2} are the mean motions of the resonant planets, with indexes j_1 and j_2.

10.1.2 The indirect effect

The indirect effect is specific to the *restricted* problem, where the motion of the planets is assumed to be a known function of time and only the motion of a massless asteroid is computed from the Hamiltonian equations. Conversely, in the planetary problem, where the motion of all the bodies is computed in a self-consistent way starting from the global Hamiltonian (2.39), we do not need to worry about the indirect effect.

In the previous parts of this book, in the study of an asteroid's dynamics we have always assumed that the planets move on fixed Keplerian orbits, or on orbits slowly precessing according to the Lagrange–Laplace solution (7.10); in both cases, the mean longitudes of the planets have been considered to be linear functions of time and all the other orbital elements of the planets have been assumed to have no short-period oscillations. This level of approximation is not sophisticated enough, when one comes to the study of three-body resonances.

In reality, the orbital elements of all planets have short periodic oscillations due to their mutual perturbations. The explicit expressions of these oscilla-

tions are obtained while constructing the secular normal form of the planetary problem, as explained in Section 2.5.

The indirect effect consists in the perturbations of the asteroid's motion that are exerted by the short-period oscillations of the orbital elements of the planets. The Hamiltonian that takes into account this effect can be obtained by substituting in the perturbation function the osculating elements of the planets with their expressions as functions of the *mean* elements. The *mean* elements are, by definition, free of short-period oscillations, and the *mean* mean longitudes are linear functions of time. Therefore, one obtains an expression that is formally equivalent to (2.38), but with $\lambda_1, \ldots, \lambda_N$ standing for the planets' *mean* mean longitudes; however, the resulting perturbation $\mathcal{H}_1$ can be written no longer as the sum of the perturbations exerted by the individual planets. In fact, in the substitution, one introduces harmonics that couple the *mean* elements of the planets, with coefficients that are quadratic in the planetary masses. Among these harmonics there are also those related to any given three-body resonance. For practical purposes, it is often enough to consider only the short-period oscillations of the orbital elements of the main perturber, i.e. Jupiter for the asteroid belt and Neptune for the Kuiper belt.

In the following, we detail the procedure for substituting the planets' osculating elements with their functions of the *mean* elements, following the approach of Nesvorný and Morbidelli (1999). In order to be explicit and constructive, this approach makes use of series expansions.

Without loss of generality, we consider only the term $\mathcal{H}_1^{(\bar{\jmath})}$ in the perturbation function (10.4), related to the $\bar{\jmath}$-th planet. In the following we restrict for simplicity to the planar case, where the inclinations of the asteroid and of the planets are set equal to 0. The substitution of the planet's orbital elements in the terms of the perturbation that depend on the inclinations can be done in an analogous way.

The classical Legendre expansion of $\varepsilon_{\bar{\jmath}}\mathcal{H}_1^{(\bar{\jmath})}$ can be written as

$$\varepsilon_{\bar{\jmath}}\mathcal{H}_1^{(\bar{\jmath})} = \frac{\varepsilon_{\bar{\jmath}}}{a_{\bar{\jmath}}} \sum_u P_u(\alpha) e^P e_{\bar{\jmath}}^{P_{\bar{\jmath}}} \cos \Psi_u \ , \tag{10.12}$$

where $\Psi_u = k_{\bar{\jmath}}\lambda_{\bar{\jmath}} + k\lambda + l_{\bar{\jmath}}\varpi_{\bar{\jmath}} + l\varpi$ (the multi-index u denoting different values of the integers $k_{\bar{\jmath}}, k, l_{\bar{\jmath}}, l$) and $\alpha = a/a_{\bar{\jmath}}$. In the expression above a, e, λ, ϖ are respectively the semimajor axis, the eccentricity, the mean longitude and the perihelion longitude of the asteroid. The variables with subscript $\bar{\jmath}$ refer to the considered planet. For the D'Alembert rules one has $k_{\bar{\jmath}} + k + l_{\bar{\jmath}} + l = 0$, and $P = |l| + 2m$ and $P_{\bar{\jmath}} = |l_{\bar{\jmath}}| + 2n$, where m, n are positive integer numbers. The functions $P_u(\alpha)$ in (10.12) can be evaluated following Šidlichovský (1990).

The expansion (10.12) contains the osculating elements of the considered planet: $a_{\bar{\jmath}}$, $\lambda_{\bar{\jmath}}$, $e_{\bar{\jmath}}$ and $\varpi_{\bar{\jmath}}$. The dependence of these variables on the *mean*

elements of the planetary system can be written in the form (see Bretagnon, 1990):

$$a_{\bar{\jmath}}=\bar{a}_{\bar{\jmath}}+\sum_v A_v\cos\Phi_v+O(\varepsilon^2) \tag{10.13}$$

$$\lambda_{\bar{\jmath}}=\bar{\lambda}_{\bar{\jmath}}+\sum_v B_v\sin\Phi_v+O(\varepsilon^2) \tag{10.14}$$

$$e_{\bar{\jmath}}\exp(\iota\varpi_{\bar{\jmath}})=\bar{e}_{\bar{\jmath}}\exp(\iota\bar{\varpi}_{\bar{\jmath}})+\sum_v C_v\exp(\iota\Phi_v)+O(\varepsilon^2)\ , \tag{10.15}$$

where $\Phi_v = r_{\bar{\jmath}}\bar{\lambda}_{\bar{\jmath}} + s_{\bar{\jmath}}\bar{\varpi}_{\bar{\jmath}} + r_j\bar{\lambda}_j + s_j\bar{\varpi}_j$ (the multi-index v denoting different values of the integers $r_{\bar{\jmath}}, r_j, s_{\bar{\jmath}}, s_j$ and of the index $j \neq \bar{\jmath}$). In the expression above, $\bar{a}_j, \bar{e}_j, \bar{\lambda}_j$ and $\bar{\varpi}_j$ denote the *mean* semimajor axis, eccentricity, mean longitude and perihelion longitude of the j-th planet; recall that the *mean* semimajor axis is constant, the *mean* mean longitude is a linear function of time, and the *mean* perihelion longitude has only secular variations. The constants A_v, B_v, C_v are proportional to the mass ε_j of the j-th planet, and $O(\varepsilon^2)$ denotes that the neglected terms are at least quadratic in the planetary masses. In a nonplanar approximation, the dependence of $i_{\bar{\jmath}}\exp(\iota\Omega_{\bar{\jmath}})$ on the planetary *mean* elements would have the same form as (10.15).

The expressions (10.13)–(10.15) can be introduced in equation (10.12). This will make $\mathcal{H}_1^{(\bar{\jmath})}$ a function of the *mean* elements. However, certain simplifications must be performed on the perturbing function in order to write it as an explicit Fourier series of cosine terms. It is suitable to treat separately the substitutions of expressions (10.13), (10.14) and (10.15) because they involve different technical procedures.

$a_{\bar{\jmath}}$ substitution

The substitution of equation (10.13) into equation (10.12) is done by first writing $P(\alpha)$ as a polynomial in α:

$$P(\alpha)=\sum_n P_n\alpha^n \tag{10.16}$$

and then substituting equation (10.13) into each term:

$$\begin{aligned}\frac{1}{a_{\bar{\jmath}}}\alpha^n&=a^n\left(\bar{a}_{\bar{\jmath}}+\sum_v A_v\cos\Phi_v\right)^{-(n+1)}\\&\cong\frac{1}{\bar{a}_{\bar{\jmath}}}\bar{\alpha}^n-(n+1)\frac{1}{\bar{a}_{\bar{\jmath}}^2}\bar{\alpha}^n\sum_v A_v\cos\Phi_v\ ,\end{aligned} \tag{10.17}$$

where we have denoted $\bar{\alpha} = a/\bar{a}_{\jmath}$, and a is the asteroid's semimajor axis. All the terms in the Taylor expansion which are of order A_v^2 or smaller have been neglected, because these terms would be at least quadratic in the planetary masses.

Substituting (10.17) in (10.16), one gets:

$$\frac{1}{a_{\jmath}} P(\alpha) = \frac{1}{\bar{a}_{\jmath}} \sum_n P_n \bar{\alpha}^n - \frac{1}{\bar{a}_{\jmath}^2} \sum_n P_n (n+1) \bar{\alpha}^n \sum_v A_v \cos \Phi_v \ . \tag{10.18}$$

Substituting (10.18) in (10.12), the first term of (10.18) recovers the original Taylor expansion of the perturbing function, now with the *mean* semimajor axis $\bar{a}_{\jmath}$, instead of the osculating one $a_{\jmath}$. The second term generates new harmonics $\cos \Psi_u \cos \Phi_v = \frac{1}{2}(\cos(\Psi_u + \Phi_v) + \cos(\Psi_u - \Phi_v))$, with coefficient:

$$-\frac{\varepsilon_{\jmath}}{\bar{a}_{\jmath}^2} A_v \left(\sum_n P_n (n+1) \bar{\alpha}^n \right) e^P e_{\jmath}^{P_{\jmath}} \ , \tag{10.19}$$

which is quadratic in the planetary masses.

$\lambda_{\jmath}$ substitution

The substitution of equation (10.14) into equation (10.12) is done by writing

$$\cos \Psi_u = \mathrm{Re} \exp \left[\iota (\bar{\Psi}_u + \sum_v k_{\jmath} B_v \sin \Phi_v)\right] \ ,$$

where Re denotes the real part, and $\bar{\Psi}_u = k_{\jmath} \bar{\lambda}_{\jmath} + k\lambda + l_{\jmath} \varpi_{\jmath} + l\varpi$. Because of

$$\exp \left(\iota k_{\jmath} B_v \sin \Phi_v\right) = \sum_n J_n(k_{\jmath} B_v) \exp\left(\iota n \Phi_v\right)$$

with coefficients given by the Bessel functions:

$$\begin{aligned} J_n(z) &= \left(\frac{z}{2}\right)^n \sum_{r \geq 0} \frac{(-1)^r}{r!(r+n)!} \left(\frac{z}{2}\right)^{2r} \ ; \quad n \geq 0 \ , \\ J_n(z) &= (-1)^n J_{-n}(z) \ ; \quad n < 0 \ , \end{aligned}$$

one can write

$$\exp \left[\iota (\bar{\Psi}_u + \sum_v k_{\jmath} B_v \sin \Phi_v)\right] = \exp\left(\iota \bar{\Psi}_u\right) \prod_v \left(\sum_n J_n(k_{\jmath} B_v) \exp\left(\iota n \Phi_v\right) \right) \ . \tag{10.20}$$

As the coefficients B_v are proportional to the planetary masses, one can approximate (10.20) retaining only the linear terms in B_v; this preserves only the terms in

$$J_0 \cong 1 \ , \quad J_1(k_{\bar{\jmath}} B_v) \cong \left(\frac{k_{\bar{\jmath}} B_v}{2}\right) \ , \quad J_{-1}(k_{\bar{\jmath}} B_v) \cong \left(-\frac{k_{\bar{\jmath}} B_v}{2}\right) \ .$$

It follows that

$$\exp\left[\iota(\bar{\Psi}_u + \textstyle\sum_v k_{\bar{\jmath}} B_v \sin\Phi_v)\right] = \exp\left(\iota\bar{\Psi}_u\right) + \frac{1}{2}\sum_v k_{\bar{\jmath}} B_v \exp\left[\iota(\bar{\Psi}_u + \Phi_v)\right] - \frac{1}{2}\sum_v k_{\bar{\jmath}} B_v \exp\left[\iota(\bar{\Psi}_u - \Phi_v)\right] \ . \quad (10.21)$$

We now substitute $\cos\Psi_u$ in (10.12) with the real part of (10.21): the first term on the right hand side of (10.21) gives the harmonic $\cos\bar{\Psi}_u$, i.e. the original harmonic with the formal substitution of the osculating $\lambda_{\bar{\jmath}}$ with the *mean* mean longitude $\bar{\lambda}_{\bar{\jmath}}$; the two other terms produce harmonics $\cos(\bar{\Psi}_u + \Phi_v)$ and $\cos(\bar{\Psi}_u - \Phi_v)$, with coefficients

$$\pm \frac{\varepsilon_{\bar{\jmath}}}{2a_{\bar{\jmath}}} k_{\bar{\jmath}} B_v P(\alpha) e^P e_{\bar{\jmath}}^{P_{\bar{\jmath}}} \quad (10.22)$$

(the upper sign for $\bar{\Psi}_u + \Phi_v$, the lower for $\bar{\Psi}_u - \Phi_v$), which are quadratic in the planetary masses.

$e_{\bar{\jmath}} \exp \iota\varpi_{\bar{\jmath}}$ substitution

We start by writing in exponential form (supposing $l_{\bar{\jmath}} \geq 0$):

$$e_{\bar{\jmath}}^{2n} e_{\bar{\jmath}}^{l_{\bar{\jmath}}} \exp\left(\iota l_{\bar{\jmath}} \varpi_{\bar{\jmath}}\right) = \left[\bar{e}_{\bar{\jmath}} \exp\left(\iota\bar{\varpi}_{\bar{\jmath}}\right) + \sum_v C_v \exp\left(\iota\Phi_v\right)\right]^n \times \left[\bar{e}_{\bar{\jmath}} \exp\left(-\iota\bar{\varpi}_{\bar{\jmath}}\right) + \sum_v C_v \exp\left(-\iota\Phi_v\right)\right]^n \times \left[\bar{e}_{\bar{\jmath}} \exp\left(\iota\bar{\varpi}_{\bar{\jmath}}\right) + \sum_v C_v \exp\left(\iota\Phi_v\right)\right]^{l_{\bar{\jmath}}}$$

Performing the binomial expansion of each factor and considering only the terms which are linear in C_v, it follows that

$$e_{\bar{\jmath}}^{2n} e_{\bar{\jmath}}^{l_{\bar{\jmath}}} \exp\left(\iota l_{\bar{\jmath}} \varpi_{\bar{\jmath}}\right) = \bar{e}_{\bar{\jmath}}^{2n} \bar{e}_{\bar{\jmath}}^{l_{\bar{\jmath}}} \exp\left(\iota l_{\bar{\jmath}} \bar{\varpi}_{\bar{\jmath}}\right) + \bar{e}_{\bar{\jmath}}^{2n} \bar{e}_{\bar{\jmath}}^{l_{\bar{\jmath}}} \exp\left(\iota l_{\bar{\jmath}} \bar{\varpi}_{\bar{\jmath}}\right) \left[(n + l_{\bar{\jmath}}) \bar{e}_{\bar{\jmath}}^{-1} \exp\left(-\iota\bar{\varpi}_{\bar{\jmath}}\right) \sum_v C_v \exp\left(\iota\Phi_v\right) + n \bar{e}_{\bar{\jmath}}^{-1} \exp\left(\iota\bar{\varpi}_{\bar{\jmath}}\right) \sum_v C_v \exp\left(-\iota\Phi_v\right)\right] \ . \quad (10.23)$$

Similarly we have:

$$
\begin{aligned}
e_{\jmath}^{2n} e_{\jmath}^{l_{\jmath}} \exp\,(-\imath l_{\jmath}\varpi_{\jmath}) =& \bar{e}_{\jmath}^{2n} \bar{e}_{\jmath}^{l_{\jmath}} \exp\,(-\imath l_{\jmath}\bar{\varpi}_{\jmath}) \\
&+ \bar{e}_{\jmath}^{2n} \bar{e}_{\jmath}^{l_{\jmath}} \exp\,(-\imath l_{\jmath}\bar{\varpi}_{\jmath}) \Bigg[(n + l_{\jmath}) \bar{e}_{\jmath}^{-1} \exp\,(\imath \bar{\varpi}_{\jmath}) \sum_v C_v \exp\,(-\imath \Phi_v) \\
&+ n \bar{e}_{\jmath}^{-1} \exp\,(-\imath \bar{\varpi}_{\jmath}) \sum_v C_v \exp\,(\imath \Phi_v) \Bigg] \ . \qquad (10.24)
\end{aligned}
$$

Thus, substituting for (10.23) and (10.24) in the term $e_{\jmath}^{P_{\jmath}} \cos \Psi_u$ of (10.12), we obtain $\bar{e}_{\jmath}^{P_{\jmath}} \cos \bar{\Psi}_u$, with $\bar{\Psi}_u = k_{\jmath}\lambda_{\jmath} + k\lambda + l_{\jmath}\bar{\varpi}_{\jmath} + l\varpi$ (i.e. the same term, rewritten with the formal substitution of $e_{\jmath}$ and $\varpi_{\jmath}$ with $\bar{e}_{\jmath}$ and $\bar{\varpi}_{\jmath}$), together with new harmonics in $\bar{\Psi}_u + \Phi_v - \bar{\varpi}_{\jmath}$ and $\bar{\Psi}_u - \Phi_v + \bar{\varpi}_{\jmath}$. The coefficients of these harmonics are respectively:

$$
(n + |l_{\jmath}|) \frac{\varepsilon_{\jmath}}{a_{\jmath}} C_v P(\alpha) e^P \bar{e}_{\jmath}^{P_{\jmath}-1} \ \text{and} \ n \frac{\varepsilon_{\jmath}}{a_{\jmath}} C_v P(\alpha) e^P \bar{e}_{\jmath}^{P_{\jmath}-1} \ . \qquad (10.25)
$$

These coefficients are, again, quadratic in the planetary masses.

10.1.3 Inclusion of both direct and indirect effects in the asteroid problem

We now show how the indirect and the direct terms can both be taken into account in the construction of the resonant normal form for a three-body resonance. First, using the recipe described in Section 10.1.2 and then introducing the modified Delaunay variables $(\Lambda, P, Q, \lambda, p, q)$ for the asteroid we write the perturbation of the restricted problem as

$$
\begin{aligned}
\varepsilon \mathcal{H}_1 =& \sum_{j=1}^{N} \varepsilon_j \mathcal{H}_1^{(j)}(\Lambda, P, Q, \lambda, p, q, \bar{\lambda}_j, \bar{\varpi}_j, \bar{\Omega}_j) \qquad (10.26) \\
&+ \sum_{j=1}^{N} \sum_{k \neq j} \varepsilon_j \varepsilon_k \mathcal{H}_2^{(j,k)}(\Lambda, P, Q, \lambda, p, q, \bar{\lambda}_j, \bar{\varpi}_j, \bar{\Omega}_j, \bar{\lambda}_k, \bar{\varpi}_k, \bar{\Omega}_k) + O(\varepsilon^3) \ ;
\end{aligned}
$$

the functions $\mathcal{H}_1^{(j)}$ are the same as in (10.4), with the formal substitution of $\lambda_j, \varpi_j, \Omega_j$ with $\bar{\lambda}_j, \bar{\varpi}_j, \bar{\Omega}_j$, while each function $\mathcal{H}_2^{(j,k)}$ depends only on the *mean* modified Delaunay variables of the j-th and of the k-th planet. The planets' *mean* mean longitudes $\bar{\lambda}_j$ are linear functions of time, with frequencies n_j. For simplicity of notation, hereafter we will omit the over-bar, and denote the *mean* mean longitudes simply by λ_j.

Then, as usual, we write the Hamiltonian of the asteroid problem in an autonomous form introducing actions Λ_j conjugate to λ_j and adding the term $\sum_j n_j\Lambda_j$ to the integrable part of the Hamiltonian. The Hamiltonian of the asteroid problem then becomes formally identical to (10.3), with the sole addition of the term $\sum_j \sum_{k\neq j} \varepsilon_j\varepsilon_k\mathcal{H}_2^{(j,k)}$, quadratic in the planetary masses, and of higher-order terms.

The averaging of the Hamiltonian over the fast angles λ, λ_j at first order in ε is then done as in Section 10.1.1, obtaining, to second order in the planetary masses, the function:

$$\varepsilon^2\mathcal{H}_2 = \frac{\varepsilon^2}{2}\Big(\{\mathcal{H}_1,\chi\} + \{\overline{\mathcal{H}}_1,\chi\}\Big) + \sum_{j=1}^{N}\sum_{k\neq j}\varepsilon_j\varepsilon_k\mathcal{H}_2^{(j,k)} \ , \tag{10.27}$$

where the first two terms on the right hand side are the same as in (10.8). This means that the perturbation terms in the averaged Hamiltonian at order two in the planetary masses are simply the sum of those coming from the direct effect and of those coming from the indirect effect. The construction of the three-body resonant normal form can then be done as explained at the end of Section10.1.1.

10.2 Three-body resonant multiplets

Once the three-body resonant normal form is computed, we can proceed as done in Section 9.3 for two-body mean motion resonances. We take into account the secular precession of planetary orbits, assuming that the mean planetary elements evolve according to the Lagrange–Laplace solution (7.10), and we introduce the angles $\varpi_j^* = g_jt+\beta_j$ and $\Omega_j^* = s_jt+\delta_j$. To make the Hamiltonian autonomous, we introduce Λ_{g_j} and Λ_{s_j} as the actions conjugate to ϖ_j^*, Ω_j^* $(j = 1,\ldots,N)$, and add to the Hamiltonian the term $\sum_j(g_j\Lambda_{g_j} + s_j\Lambda_{s_j})$.

The normal form for a three-body resonance $k_1\dot\lambda_{j_1} + k_2\dot\lambda_{j_2} + k\dot\lambda \sim 0$ is therefore written as:

$$\begin{aligned}\mathcal{H}_{3\mathrm{BR}} =& -\frac{1}{2\Lambda^2} + n_{j_1}\Lambda_{j_1} + n_{j_2}\Lambda_{j_2} + \sum_{j=1}^{N}(g_j\Lambda_{g_j} + s_j\Lambda_{s_j}) + \varepsilon\bar{\mathcal{H}}_1(p,q,\boldsymbol{\varpi}^*,\boldsymbol{\Omega}^*)\\ &+\varepsilon^2\mathcal{H}_2(k_1\lambda_{j_1} + k_2\lambda_{j_2} + k\lambda,\boldsymbol{\varpi}^*,\boldsymbol{\Omega}^*,p,q) \ ,\end{aligned} \tag{10.28}$$

where $n_{j_1} = \dot\lambda_{j_1}, n_{j_2} = \dot\lambda_{j_2}$ and, as in previous chapters, $\boldsymbol{\varpi}^*$ and $\boldsymbol{\Omega}^*$ denote the vectors with components ϖ_j^* and Ω_j^* respectively. Both functions $\bar{\mathcal{H}}_1$ and $\mathcal{H}_2$ in (10.28) depend on the actions Λ, P, Q of the asteroid, and $\mathcal{H}_2$ can be

Table 10.1: Coefficients for the 5 −2 −2 three-body resonance. Reprinted from Table 3 of Nesvorný and Morbidelli (1999), with permission from Kluwer Academic Publishers.

N.	Perihelia	$a_{\rm res} - 3.17$ (10^{-3} AU)	$c_{u_5,u_6,-s}$ (10^{-8})
1	$-\varpi$	5.08	$45.59e - 32.24e^3$
2	$-\varpi_5^*$	4.33	$-2.76 + 0.93e^2$
3	$-\varpi_6^*$	4.44	$1.18 - 0.38e^2$
4	$\varpi_5^* - \varpi_6^* - \varpi$	5.19	$-3.82e + 2.07e^3$
5	$-\varpi_5^* + \varpi_6^* - \varpi$	4.97	$-7.39e - 2.98e^3$
6	$-2\varpi_5^* + \varpi$	3.57	$0.15e + 0.73e^3$
7	$-2\varpi_6^* + \varpi$	3.79	$0.03e + 0.14e^3$
8	$-\varpi_5^* - \varpi_6^* + \varpi$	3.68	$-0.13e - 0.60e^3$
9	$-2\varpi_5^* + \varpi_6^*$	4.21	$0.01 + 0.07e^2$
10	$\varpi_5^* - 2\varpi_6^*$	4.55	0.01
11	$\varpi_5^* - 2\varpi$	5.83	$-23.5e^2$
12	$\varpi_6^* - 2\varpi$	5.72	$23.2e^2$
13	$2\varpi_5^* - 3\varpi$	6.58	$0.64e^3$
14	$2\varpi_6^* - 3\varpi$	6.36	$1.42e^3$
15	$\varpi_5^* + \varpi_6^* - 3\varpi$	6.47	$3.41e^3$

expanded in a Fourier series as

$$\mathcal{H}_2 = \sum_{m,\mathbf{u},\mathbf{v},s,r} c_{m,\mathbf{u},\mathbf{v},s,r}(\Lambda, P, Q) \exp\left(\iota\sigma_{m,\mathbf{u},\mathbf{v},s,r}\right) \tag{10.29}$$

where $\sigma_{m,\mathbf{u},\mathbf{v},s,r} = m(k_1\lambda_{j_1} + k_2\lambda_{j_2} + k\lambda) + \mathbf{u}\cdot\boldsymbol{\varpi}^* + \mathbf{v}\cdot\boldsymbol{\Omega}^* + sp + rq$ and m, r, s are integer numbers and $\mathbf{u} \equiv (u_1, \ldots, u_n)$ and $\mathbf{v} \equiv (v_1, \ldots, v_n)$ are vectors with integer components such that $m(k + k_1 + k_2) - s - r + \sum_j(u_j + v_j) = 0$.

As explained in Section 9.3, each $\sigma_{m,\mathbf{u},\mathbf{v},s,r}$ is an independent resonant angle, so that a three-body resonance splits into a multiplet of resonances. Because the coefficients of the resonant harmonics are of order ε^2, the resonances of the multiplet should have a width in the semimajor axis of order ε; the separation in semimajor axis of the different resonances, being proportional to the precession frequencies of the secular angles, should also be of order ε. Therefore we expect that in general a three-body resonance is characterized by a marginal overlap of its resonant multiplet. The paper by Nesvorný and Morbidelli (1999) explicitly computes the multiplet structure of some notable three-body resonances occurring in the asteroid belt.

Table 10.1 lists the main harmonics of the multiplet of the resonance $5\dot\lambda_{\rm J} - 2\dot\lambda_{\rm S} - 2\dot\lambda \sim 0$ (hereafter 5 −2 −2 three-body resonance), where $\dot\lambda_{\rm J}$ and $\dot\lambda_{\rm S}$ are

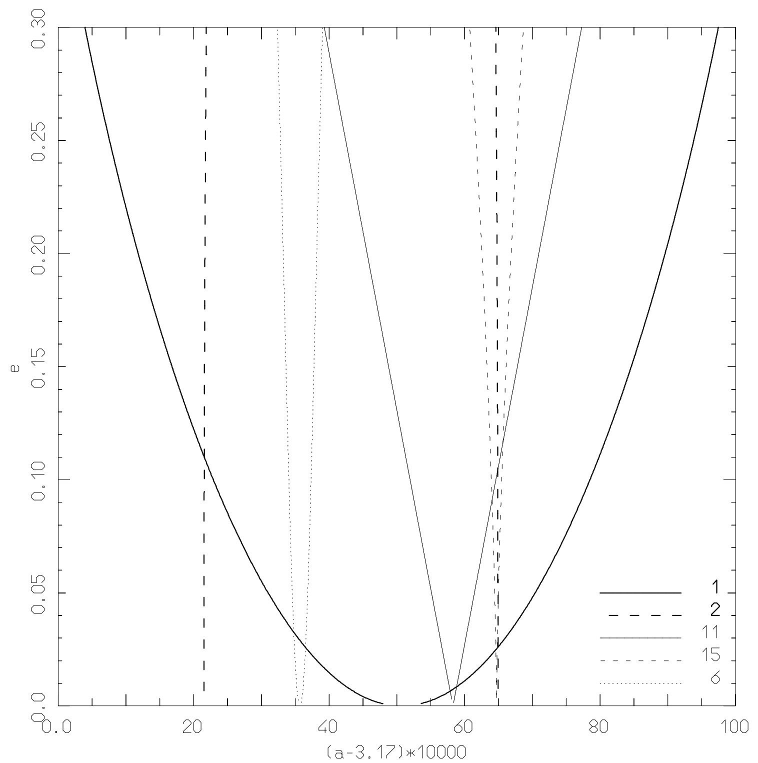


Figure 10.1: Separatrices of five multiplet resonances of the 5 −2 −2 three-body resonance. Reprinted from Fig. 2 of Nesvorný and Morbidelli (1999), with permission from Kluwer Academic Publishers.

the *mean* mean motions of Jupiter and Saturn and $\dot{\lambda}$ is the mean motion of the asteroid, respectively. This resonance occurs at about 3.17 AU in the asteroid belt. The first column gives the reference number of the harmonic. The second one shows the combination of the perihelia $u_5\varpi_5^* + u_6\varpi_6^* - s\varpi$ that characterizes the resonant angle $\sigma_{1,(0,\ldots,0,u_5,u_6,0\ldots),\mathbf{0},s,0}$ (hereafter denoted for simplicity as $\sigma_{u_5,u_6,-s}$). The third column shows the value of the semimajor axis $a_{\rm res}$ at which the angle $\sigma_{u_5,u_6,-s}$ is exactly resonant, taking into account the precession frequencies of the perihelia. Finally, the last column lists the coefficient of the harmonic $c_{1,(0,\ldots,0,u_5,u_6,0\ldots),\mathbf{0},s,0}$ (hereafter $c_{u_5,u_6,-s}$) and its dependence on the asteroid's eccentricity. The computations have been done in the framework of the planar problem, so that all terms involving the longitudes of nodes and the inclinations are not reported.

As one sees from Table 10.1, the dominant harmonic in the multiplet is the first one. The resonant angle is $\sigma_{0,0,-1} = 5\lambda_{\rm J} - 2\lambda_{\rm S} - 2\lambda - \varpi$ and the linear

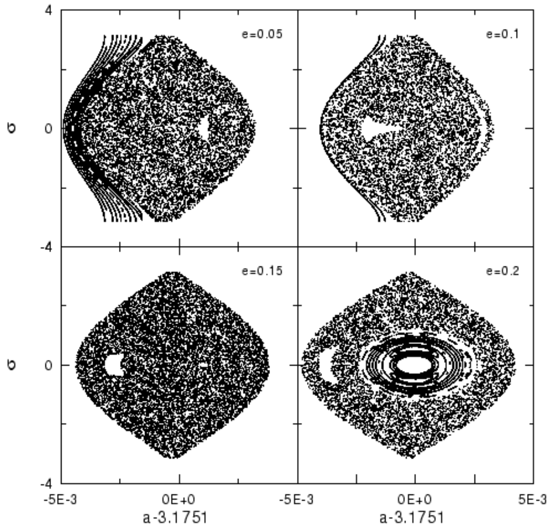


Figure 10.2: Poincaré sections of the model (10.30) for the 5 −2 −2 three-body resonance. The semimajor axis and $\sigma = \sigma_{0,0,-1}$ are plotted when $\varpi - \varpi_5^* = 0$. Reprinted from Fig. 3 of Nesvorný and Morbidelli (1999), with permission from Kluwer Academic Publishers.

part in the eccentricity of the coefficient of the harmonic is $45.59 \times 10^{-8} e$. This harmonic is mainly generated by the $e_{\rm J} \exp(\iota\varpi_{\rm J})$ contribution of the indirect effect and, more precisely, the large value of its coefficient is due to the favorable combination of the term $\Phi_v = \bar{\lambda}_{\rm J} - 2\bar{\lambda}_{\rm S} + \bar{\varpi}_{\rm J}$ in (10.15) with the term $\Psi_u = 4\lambda_{\rm J} - 2\lambda - \varpi_{\rm J} - \varpi$ in (10.12).

The structure of the resonant multiplet of the 5 −2 −2 resonance is complex. Looking carefully at Table 10.1, one realizes that the multiplet is made of submultiplets of resonances which almost completely overlap. For instance resonances 4 and 5 overlap resonance 1 (they have the same dependence on the asteroid's eccentricity and almost exactly the same resonant semimajor axis). Similarly, resonances 3 and 9 overlap resonance 2; resonance 12 overlaps resonance 11; resonances 13 and 14 overlap resonance 15 and resonances 7 and 8 overlap resonance 6. Recall from Section 9.4 that, when the separation among the resonances is much smaller than their width, the resulting dynamics is similar to that of a modulated pendulum. Therefore, each of these submultiplets should be considered as a modulated pendulum-like resonance.

In addition, these modulated pendulum-like resonances overlap each other, as shown in Fig. 10.1, where the separatrices of the resonances 1, 2, 11, 15 and 6 in Table 10.1 (namely the dominating terms of each submultiplet) are plotted.

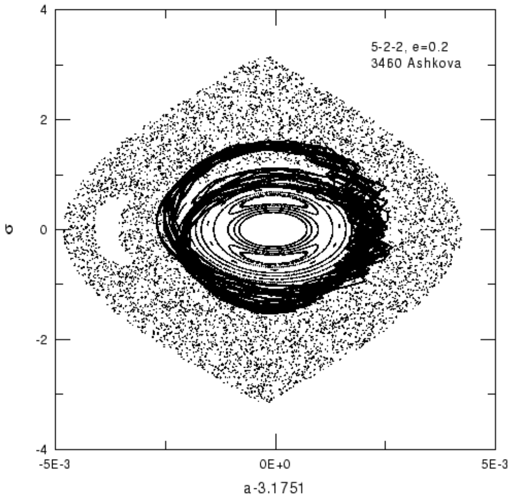


Figure 10.3: A composite of the Poincaré section of the 5 −2 −2 three-body resonance at $e = 0.2$ and of the filtered motion of 3460 Ashkova. Reprinted from Fig. 4 of Nesvorný and Morbidelli (1999), with permission from Kluwer Academic Publishers.

This overlap should be considered to be responsible for the general global chaoticity of the 5 −2 −2 three-body resonance. Only at large eccentricity, where resonance 1 is much larger than all other resonances, should a central island of stable motion be expected, as in a modulated pendulum of type (9.20), in the case where the relative amplitude of modulation $\alpha < 1$.

An approximate view of the dynamics generated by this multiplet structure can be obtained by computing Poincaré surfaces of section for a simplified model that includes only the two major harmonics reported in Table 10.1, given by the Hamiltonian:

$$\mathcal{H} = -\frac{1}{2\Lambda^2} + n_{\rm J}\Lambda_{\rm J} + n_{\rm S}\Lambda_{\rm S} + g_5\Lambda_{g_5} + \beta e^2 + c_{0,0,-1}\cos\sigma_{0,0,-1} + c_{-1,0,0}\cos\sigma_{-1,0,0}\ . \tag{10.30}$$

This Hamiltonian is in fact the simplified version of (10.28), where only two harmonics have been taken into account in $\varepsilon^2\mathcal{H}_2$, and the secular part $\varepsilon\overline{\mathcal{H}}_1$ has been reduced to the sole term βe^2.

Figure 10.2 shows the Poincaré sections on four energy surfaces, which are identified on the upper right-hand corner of each panel by the eccentricity at which the energy surface $\mathcal{H} =$constant (with $\mathcal{H}$ as in 10.30 and the angles set equal to zero) intersects the resonant semimajor axis $a_{\rm res} = 3.1751$ AU. On such energy surfaces, the eccentricity is roughly constant over the semimajor axis range reported in the figure.

Table 10.2: The same as Table 10.1 but for the 6 1 −3 three-body resonance. Reprinted from Table 6 of Nesvorný and Morbidelli (1999), with permission from Kluwer Academic Publishers.

N.	Perihelia	$a_{\rm res} - 3.13$ (10^{-3} AU)	$c_{u_5,u_6,-s}$ (10^{-8})
1	-4ϖ	8.88	$-28.45e^4$
2	$-\varpi_5^* - 3\varpi$	8.50	$10.52e^2$
3	$-\varpi_6^* - 3\varpi$	8.57	$-16.84e^3$
4	$-2\varpi_5^* - 2\varpi$	8.13	$-1.28e^2$
5	$-2\varpi_6^* - 2\varpi$	8.27	$-1.16e^2$
6	$-\varpi_5^* - \varpi_6^* - 2\varpi$	8.20	$3.39e^2$
7	$-3\varpi_5^* - \varpi$	7.75	$0.07e - 0.20e^3$
8	$-3\varpi_6^* - \varpi$	7.96	$-0.01e^3$
9	$-2\varpi_5^* - \varpi_6^* - \varpi$	7.82	$-0.22e + 0.43e^3$
10	$-\varpi_5^* - 2\varpi_6^* - \varpi$	7.89	$0.17e - 0.42e^3$

The chaotic motion fills almost all the resonant space in the three panels corresponding to $e = 0.05$, 0.1 and 0.15. Only for $e = 0.2$ is a relatively large portion of the phase space occupied by a central regular area, as expected from the fact that, for this value of the eccentricity, the resonance in the multiplet with angle $\sigma_{0,0,-1}$ is almost twice as wide as the one with angle $\sigma_{-1,0,0}$ (see Fig. 10.1). Of course, the real dynamics is much more complicated than that shown in Fig. 10.2, because the model (10.30) is strongly simplified. However, the integration of the full equations of motion for the asteroid 3460 Ashkova confirms the existence of a quasi-regular region in this three-body resonance at $e \sim 0.2$. Figure 10.3 shows the evolution of Ashkova's trajectory over 10^5 y, filtered over the short-period oscillations, and superposed on the corresponding Poincaré section of model (10.30). Note the quite good agreement on this timescale. However Ashkova's trajectory becomes irregular on a longer timescale and its Lyapunov time is about 8300 years.

The overlapping of the resonances in the multiplet is at the origin of the chaotic behavior also for most of the other three-body resonances of low order, and several additional examples are provided by Nesvorný and Morbidelli (1999). Only resonances of higher eccentricity order can be characterized by the separation of the multiplet components at low eccentricity.

An example of this is provided by the 6 1 −3 three-body resonance, which is of order 4 in the eccentricity. Table 10.2 gives the coefficients of the harmonics of the multiplet and Fig. 10.4 shows the separatrices of the corresponding

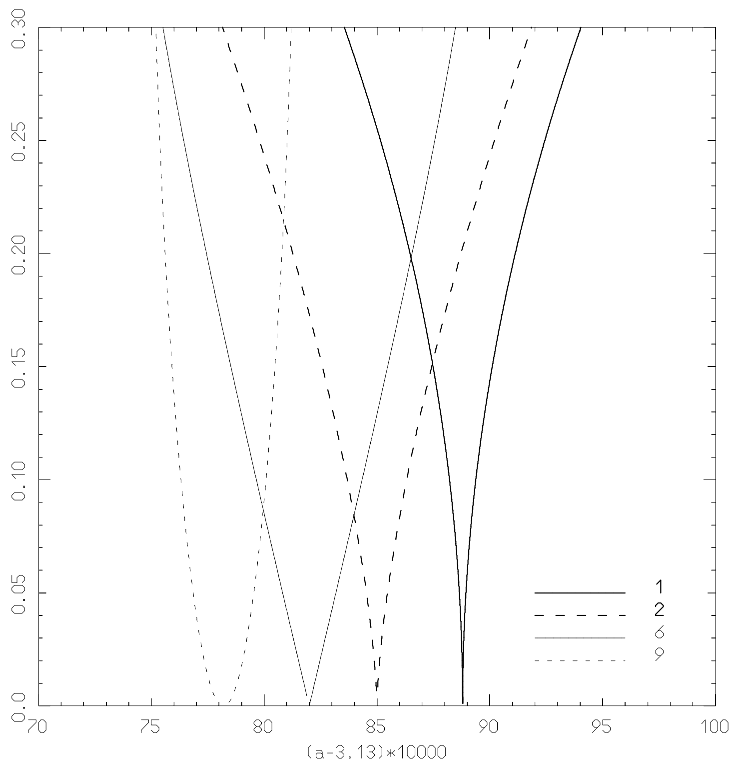


Figure 10.4: Separatrices of four multiplet resonances of the 6 1 −3 three-body resonance. Reprinted from Fig. 8 of Nesvorný and Morbidelli (1999), with permission from Kluwer Academic Publishers.

resonances: the region of nonoverlap extends to $e \sim 0.07$, a value that is of the order of the amplitude of the secular oscillations of the asteroid eccentricity forced by Jupiter.

For three-body resonances of increasingly higher order, the width of each resonant component should shrink, while their mutual distance should remain of the same order (dictated by the secular precession rates of the perihelia). Therefore the region where the resonances of the multiplet do not overlap should extend to larger eccentricities than those observed in Fig. 10.4. Consequently, the dynamics of three-body resonances of very large eccentricity order should be basically regular.

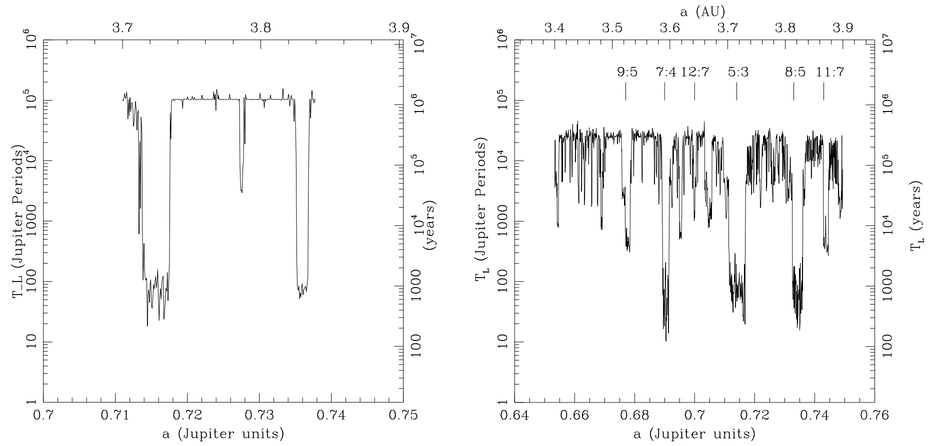


Figure 10.5: Lyapunov time as a function of the semimajor axis in the outer asteroid belt. Left panel: the model includes only the effect of Jupiter, assumed on an elliptic orbit that secularly evolves according to (7.10). Right panel: the model includes the effects of all four giant planets, fully interacting each other. The main mean motion resonances with Jupiter are labeled at the top of the right panel; all other chaotic zones are then presumably associated to three-body resonances. Reprinted from Figs. 1 and 4 of Murray *et al.* (1998), with permission from the American Astronomical Society.

10.3 The asteroid and Kuiper belts

Three-body resonances with Jupiter and Saturn are extremely important for understanding the chaotic structure of the asteroid belt. A very clear example is provided by Fig. 10.5 (from Murray *et al.*, 1998) which shows the Lyapunov time (inverse of the maximum Lyapunov exponent) as a function of the initial semimajor axis in the outer asteroid belt. The left panel is computed in a model that takes into account only the perturbations of Jupiter on a secularly precessing Keplerian orbit, while the right panel is the result of a computation which includes both Jupiter and Saturn (as well as Uranus and Neptune), fully interacting with each other.

As discussed in Section 5.2 (see Fig. 5.5), each dip and each peak in the Lyapunov time profile denotes the position of a chaotic zone, while a regular plateau, where the Lyapunov time is almost constant as a function of the initial condition, reveals a region where resonances have a negligible effect and the dynamics is expected to be essentially regular. The fact that the Lyapunov time in this regular region is not infinite is dictated by the limited integration time. Keeping these premises in mind, Fig. 10.5 reveals few isolated chaotic regions, separated by intervals of regular motion, in the model that accounts

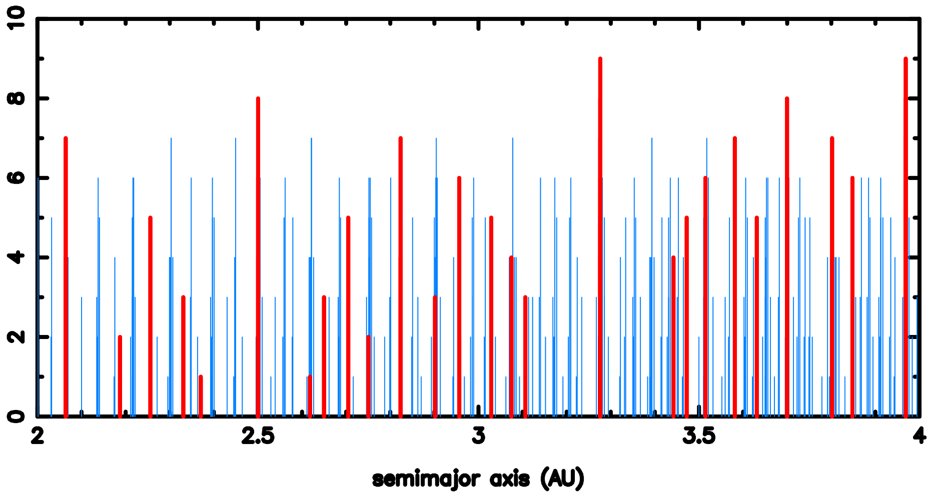


Figure 10.6: Location of the mean motion resonances $k_J\dot{\lambda}_J - k\dot{\lambda} \sim 0$ with $k_J + k \leq 20$ (bold red lines) and of the three-body resonances $k_J\dot{\lambda}_J + k_S\dot{\lambda}_S + k\dot{\lambda} \sim 0$ with $|k_J| + |k_S| + |k| \leq 20$ (light blue lines) in the asteroid belt. The vertical scale gives a qualitative comparative indication of the strength of the resonances. Adapted from Nesvorný and Morbidelli (1998).

only for the perturbations of Jupiter; conversely, in the model that includes Saturn, the dynamics is essentially chaotic over all the considered semimajor axis range. From the point of view of the resonant structure, the difference between the two models is the presence in the second one of the three-body Jupiter–Saturn–asteroid resonances.

The reason that three-body resonances change so drastically the chaotic structure of the belt is that, involving three mean motion frequencies, they are much more numerous than the ordinary mean motion resonances with Jupiter of comparable strength. To illustrate this property, Fig. 10.6 marks with vertical lines the locations of both the two-body and the three-body mean motion resonances, for semimajor axis ranging over all the asteroid belt (2–4 AU). For the two-body mean motion resonances with Jupiter (red lines) the height of the line is given as $10 - q$, where q is the eccentricity order of the resonance, so that all the resonances up to order 9 are shown. For the three-body mean motion resonances (blue lines) the height of the line is given instead as $7 - q$. This is done to account for the fact that the coefficients of the three-body resonant harmonics are quadratic in the masses and that for an eccentricity $e = 0.066$ the mass of Saturn (in solar units) is equal to e^3; in other words, a three-body resonance of a given order q should have roughly the

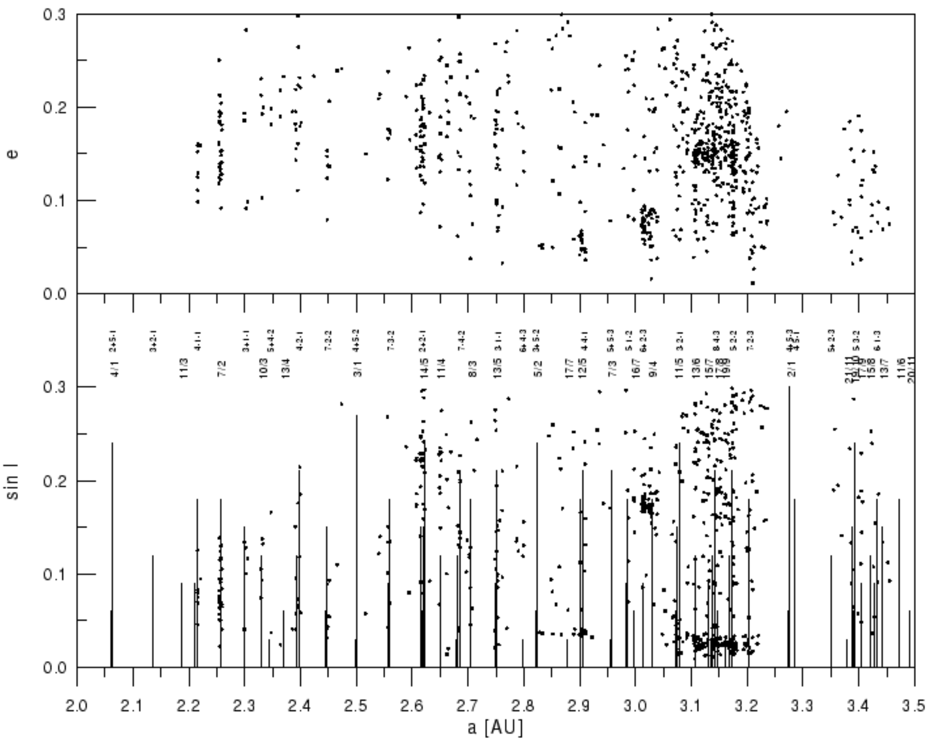


Figure 10.7: The numbered asteroids that are strongly chaotic under the effect of the giant planets, according to Froeschlé *et al.* (1997). The locations of the main two-body and three-body resonances are also indicated with vertical lines. The heights of the lines are scaled as in Fig. 10.6. Above each line the integer coefficients characterizing the resonance are reported: $k_{\rm J}$ $k_{\rm S}$ k denote the three-body resonance $k_{\rm J}\dot\lambda_{\rm J} + k_{\rm S}\dot\lambda_{\rm S} + k\dot\lambda \sim 0$, while $k_{\rm J}/k$ denotes the two-body resonance $k_{\rm J}\dot\lambda_{\rm J} - k\dot\lambda \sim 0$. Reprinted from Fig. 5 of Nesvorný and Morbidelli (1998), with permission from the American Astronomical Society.

same strength as a resonance with Jupiter of order $q + 3$ for eccentricity about 0.05–0.10. With this trick, the vertical scale of Fig. 10.6 gives a qualitative comparative indication of the strength of the various resonances, so that the overdensity of the three-body resonances appears evident.

Unlike low-order mean motion resonances with Jupiter, which are so powerful to remove the asteroids at their locations (for the reasons explained in the next chapter), the three-body resonances are not associated to any remarkable gap in the asteroids' distribution. As a consequence, most of the chaotic *real* asteroids are in three-body resonances. Figure 10.7 shows the semimajor axis,

eccentricity and inclination of 836 asteroids that have been found to be chaotic by Froeschlé *et al.* (1997), in a model taking into account the perturbations of the four giant planets (the figure makes use of the proper elements by Milani and Knežević (1994) when available, and of the osculating elements otherwise). The bottom panel shows the location of some of the main mean motion resonances with Jupiter (bold lines) and three-body Jupiter–Saturn–asteroid resonances (light lines). The height of the lines has been scaled as in Fig. 10.6 (note that with respect to Fig. 10.6, only some of the resonances have been plotted to make the picture readable).

The figure shows that most of the chaotic asteroids are concentrated along lines that can be easily identified with two-body and three-body resonances of moderate order. The two-body mean motion resonances 7/2, 10/3, 11/4, 8/3, 9/4, 13/6, etc., and the three-body mean motion resonances 4 −1 −1, 3 1 −1, 4 −2 −1, 5 3 −2, 5 2 −2, 2 2 −1, 4 −3 −1, 5 1 −2, 3 −1 −1, 6 2 −3, 5 −2 −2, 5 3 −3, etc., seem to be responsible for most of the chaotic orbits of the real asteroids in the belt. These apparent associations have been checked by Nesvorný and Morbidelli (1998) with numerical integrations covering 10^5 y, by monitoring the evolution of their filtered resonant angles. 255 asteroids have been unambiguously identified to be in three-body resonances and 63 in mean motion resonances with Jupiter. Taking into account that the original database of 836 chaotic asteroids has been obtained by Froeschlé *et al.* (1997) from the selection of the first 5400 numbered asteroids, which constitute only less than 1/6 of all the catalogued asteroids, Nesvorný and Morbidelli concluded that at least 1500 real asteroids should be chaotic due to three-body resonances and 380 due to mean motion resonances with Jupiter. Notice that these numbers must be severe low estimates, because the positive identification of a chaotic asteroid with a resonance by Nesvorný and Morbidelli (1998) was less than 50% efficient (318 asteroids with identified resonance out of 836), and because the work of Froeschlé *et al.* (1997), being based on the computation of fast Lyapunov indicators over 50,000 y (see Section 5.4.1), probably selected only the asteroids with the strongest chaotic motion. Conversely, the ratio 255/63 of the number of bodies in three-body resonances vs. two-body resonances should not be biased by the computational methods, and should therefore reflect the reality of the asteroid belt.

In Fig. 10.7, most of the chaotic asteroids appear to be concentrated between 3.1 and 3.25 AU and $e > 0.1$; moreover, unlike the other parts of the belt, in this region the chaotic asteroids are not located along well-defined lines $a \sim$ constant. A detailed exploration of this region has been done by Nesvorný and Morbidelli (1998). Figure 10.8 shows (solid line) the maximum Lyapunov exponent (MLE) as a function of the semimajor axis, computed

Table 10.3: Numerical results on three-body resonances in the asteroid belt. The first entry lists the numbers k_J, k_S, k that characterize the resonance $k_J\dot{\lambda}_J + k_S\dot{\lambda}_S + k\dot{\lambda} \sim 0$; the second column gives the identification number of the resonant asteroid whose motion, integrated over 2×10^5 y, provides the informations reported in the following columns: proper eccentricity (e_p), center value (a_c) and amplitude (δa) of its semimajor axis oscillations and the typical period T of these oscillations. The latter is typically of the order of the Lyapunov exponent. Note that, the motion of most of the asteroids being chaotic, all data reported in this table are only indicative of an order of magnitude. Reprinted from Table 1 of Nesvorný and Morbidelli (1998), with permission from the American Astronomical Society.

Resonance	Asteroid	e_p	a_c (AU)	δa (10^{-3} AU)	T (10^3 y)
4 −1 −1	2440	0.1113	2.2157	0.6	50
4 −2 −1	463	0.1795	2.3977	3.0	∼10
7 −2 −2	1966	0.1241	2.4476	0.6	∼30
7 −3 −2	1430	0.1741	2.5599	0.5	∼30
2 2 −1	258	0.1687	2.6155	0.7	∼20
6 −1 −2	53	0.2092	2.6190	∼1.0	35
4 −3 −1	792	0.1604	2.6230	2.5	25
7 −4 −2	789	0.1471	2.6857	0.5	20
3 −1 −1	485	0.1958	2.7525	3.0	15
4 −4 −1	22	0.0881	2.9095	∼1.0	∼50
5 −1 −2	576	0.1758	2.9860	2.0	20
3 −2 −1	2395	0.0690	3.0790	4.0	10
6 1 −3	936	0.1540	3.1385	0.4	10
8 −4 −3	10	0.1347	3.1418	0.5	∼30
3 3 −2	106	0.1466	3.1708	∼2.0	
5 −2 −2	490	0.0652	3.1738	4.0	10
7 −2 −3	530	0.1937	3.2080	<8.0	12

by integrating fake asteroids for 2 My under the perturbations of the four giant planets. The bottom, middle and top panels refer respectively to asteroids with initial eccentricity equal to 0.05, 0.15 and 0.25. The dotted lines denote the relative change of the asteroid's perihelion frequency over 1 My. As explained in Chapter 5, MLE and frequency analysis give complementary information: the MLE tells about the degree of chaoticity, while the relative change of perihelion frequency is related to the speed of chaotic diffusion in the frequency space. Recall that for both indicators the existence of a plateau indicates a smooth dependence of the results on the initial conditions, typical of regions where effective resonances are absent, while a peak reveals the existence of a chaotic region. As one can see in Fig. 10.8, the chaotic regions are

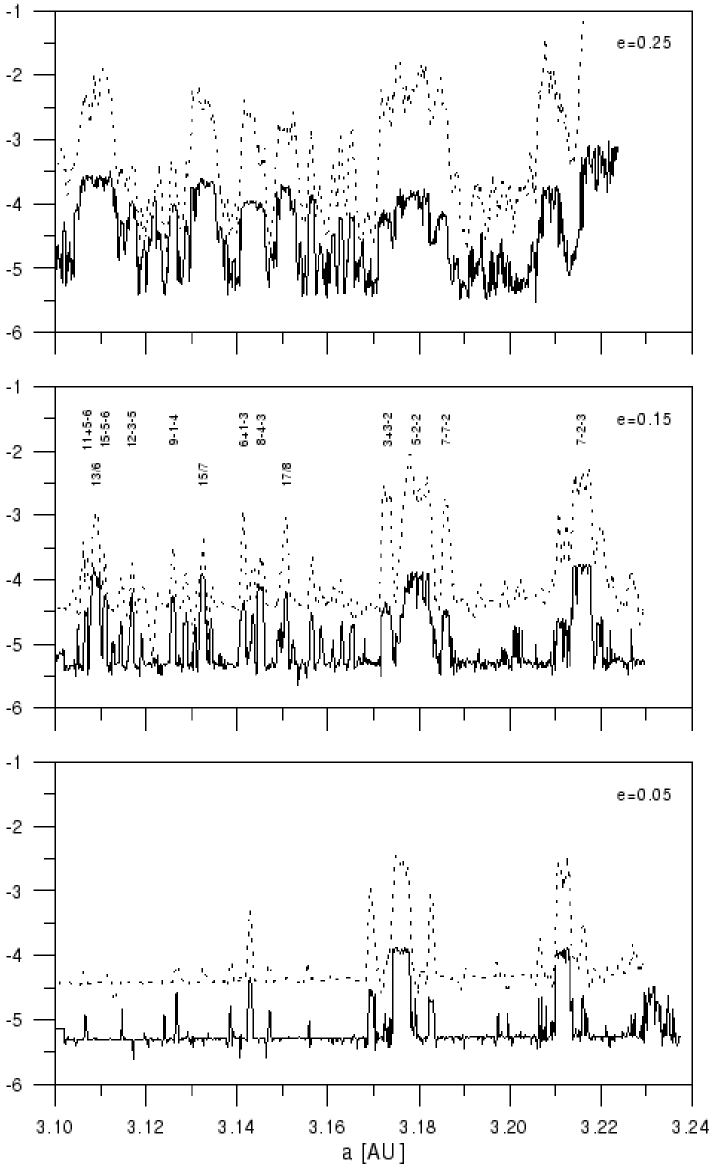


Figure 10.8: Lyapunov exponent (solid curve) and relative change of the asteroid's perihelion frequency (dotted curve) as a function of semimajor axis for three values of the eccentricity. In this region a large number of chaotic asteroids are located. Reprinted from Fig. 6 of Nesvorný and Morbidelli (1998), with permission from the American Astronomical Society.

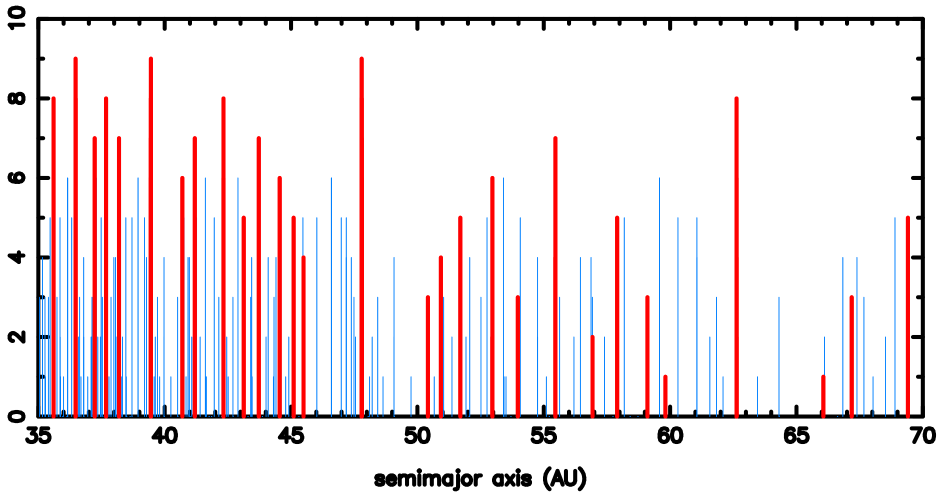


Figure 10.9: The same as Fig. 10.6, but for the mean motion resonances with Neptune (red) and the three-body resonances with Uranus and Neptune (blue) in the Kuiper belt.

well correlated with the location of high-order mean motion resonances with Jupiter and three-body Jupiter–Saturn–asteroid resonances. With increasing eccentricity the widths of the chaotic regions become larger, because all these resonances are related to harmonics whose coefficients are positive powers of e; moreover new peaks appear, related to high-order resonances that are ineffective at low e. As a consequence, the chaotic regions overlap starting from $e \sim 0.15$, making this part of the belt globally chaotic: this explains why the chaotic asteroids do not appear to be located along well-defined lines. Note that this eccentricity threshold for global resonance overlap is much lower than the one indicated in Fig. 9.12, because the work by Dermott and Murray (1983) did not take into account the resonances with Jupiter of large order and, of more importance, the three-body resonances with Jupiter and Saturn. Table 10.3 gives a compilation of the properties of the most important three-body resonances of the main asteroid belt.

In the Kuiper belt the most important three-body resonances are those with the mean motions of Uranus and Neptune. Figure 10.9 is the equivalent of Fig. 10.6, but reports the main mean motion resonances with Neptune and three-body resonances with Uranus and Neptune between 35 and 70 AU. To account for the mass of Uranus being smaller than the mass of Saturn, the height of the blue lines is now equal to $6 - q$ (instead of $7 - q$ as in Fig. 10.6),

where q is the eccentricity order of the corresponding resonance. The situation is analogous to that of the asteroid belt: the three-body resonances are much denser in the semimajor axis than the ordinary two-body resonances of comparable strength, particularly beyond 45 AU. Therefore, the three-body resonances are an essential ingredient to understanding the global dynamical structure and the long-term stability properties of the Kuiper belt, as will be shown in Chapter 12. Unfortunately, for the limited number of bodies discovered so far and for the still inaccurate determination of the orbital parameters, it is not yet possible to estimate which fraction of the real Kuiper belt population is chaotic due to three-body resonances.

10.4 Chaotic dynamics of the giant planets

As discussed in Chapter 7, the secular motion of the terrestrial planets is strongly chaotic; conversely, the secular motion of the giant planets presents only a very weak chaoticity that is raised by the coupling with the terrestrial planets. Therefore, without the presence of the terrestrial planets, the giant planets by themselves would constitute a regular system, at least from the point of view of secular dynamics.

Nevertheless, Sussman and Wisdom (1992) computed a positive Lyapunov exponent in a numerical integration of the giant planets. However this result has never been stressed, because the Lyapunov exponent presented a puzzling dependence on the integration step-size and on the initial conditions. Moreover, Sussman and Wisdom failed to identify a resonance responsible for the origin of chaos. In fact, the giant planets are not in two-body mean motion resonances, even of considerably large order. It is well known that Jupiter and Saturn are close to the 5/2 resonance, Uranus and Neptune are close to the 2/1 resonance and Jupiter and Uranus are close to the 7/1 resonance, but in all cases the critical resonant angles circulate, showing that the planets are not in such resonant configurations.

Recently, Murray and Holman (1999) showed that the system of the giant planets presents two three-body resonant relationships, which are:

$$3\dot{\lambda}_{\rm J} - 5\dot{\lambda}_{\rm S} - 7\dot{\lambda}_{\rm U} \sim 0 \ , \quad \text{and} \quad 3\dot{\lambda}_{\rm S} - 5\dot{\lambda}_{\rm U} - 7\dot{\lambda}_{\rm N} \sim 0 \qquad (10.31)$$

where $\dot{\lambda}_{\rm J}, \dot{\lambda}_{\rm S}, \dot{\lambda}_{\rm U}$ and $\dot{\lambda}_{\rm N}$ denote the mean motions of Jupiter, Saturn, Uranus and Neptune, respectively. These resonances couldn't be detected by Laskar (1985) in the computation of the secular normal form, because their resonant terms appear at order 2 in the planetary masses and order 9 in the eccentricity, which were beyond Laskar's computational limits. As estimated by Murray and Holman, the coefficients of the harmonics related to the first resonance

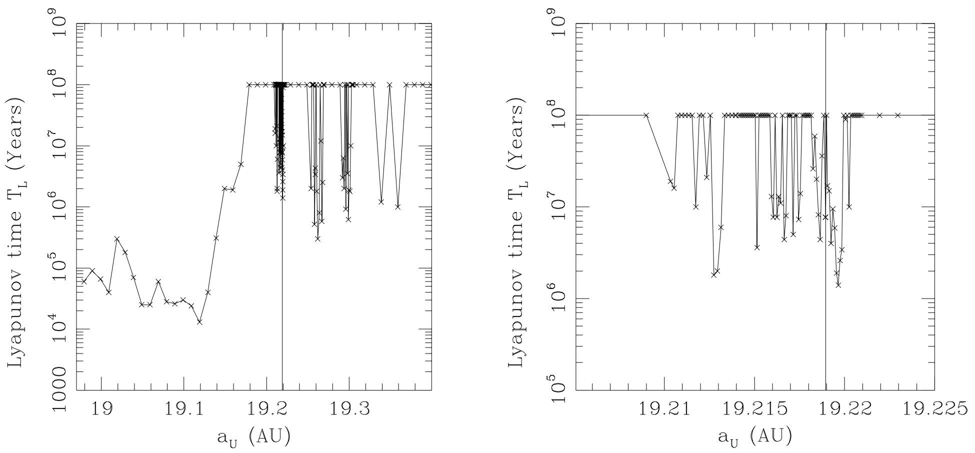


Figure 10.10: Lyapunov time as a function of the initial value of the semimajor axis of Uranus $a_{\rm U}$. See text for detailed description. Reprinted from Figs. 1 and 2 of Murray and Holman (1999), with permission from the American Association for the Advancement of Science.

in the equations of motion of Uranus are of order 10^{-12} in units such that $\mathcal{G}M_\odot = a_{\rm U} = 1$, while the coefficients of the harmonics related to the second resonance are smaller by a factor

$$\sim (\varepsilon_{\rm N}/\varepsilon_{\rm J}) * [(5\dot\lambda_{\rm S} - 2\dot\lambda_{\rm J})/(2\dot\lambda_{\rm N} - \dot\lambda_{\rm U})] \sim 3 \times 10^{-3} \ .$$

To understand in detail the complex dynamical structure of the giant planet system, Murray and Holman made a numerical survey in which all the initial orbital elements of the planets were fixed, except Uranus' semimajor axis $a_{\rm U}$, which ranged from 18.9789 and 19.3990 AU. For every set of initial conditions, the motion was integrated for 200 My, and the Lyapunov time was computed. The left panel of Fig. 10.10 shows the resulting dependence of the Lyapunov time as a function of the initial value of $a_{\rm U}$. The strong chaotic region extending up to 19.13 AU is due to the 2/1 resonance with Neptune, and overlaps with the chaotic region generated by the 1/7 resonance with Jupiter in the range 19.13–19.17 AU. Conversely, the chaotic regions centered at $a_{\rm U} = 19.219, 19.26, 19.29$ and 19.34 AU are all associated to three-body resonances. The present value of Uranus' semimajor axis, as given by the DE200 ephemeris of the Jet Propulsion Laboratory, is indicated by a vertical line in the figure. The right panel of Fig. 10.10 is a close-up around this value. For

$a_U < 19.218$ AU one can distinguish the thin chaotic regions generated by the individual resonant components of the multiplet of the three-body resonance $3\dot{\lambda}_J - 5\dot{\lambda}_S - 7\dot{\lambda}_U \sim 0$, separated by narrow regular regions. More precisely, between 19.216 and 19.218 AU Murray and Holman identified the resonant components with critical angles $3\lambda_J - 5\lambda_S - 7\lambda_U + q\varpi_5^* + 7\varpi_6^* + (2-q)\varpi_7^*$, for various values of the integer q. Conversely, for $19.218 \leq a_U \leq 19.22$ AU, the individual resonant components partially overlap, so that nearly all the orbits are chaotic and have a finite Lyapunov time. The critical angles of these resonant components are $3\lambda_J - 5\lambda_S - 7\lambda_U + q\varpi_5^* + 6\varpi_6^* + (3-q)\varpi_7^*$, but also some resonant components of the three-body resonance $3\dot{\lambda}_S - 5\dot{\lambda}_U - 7\dot{\lambda}_N \sim 0$ are present, more precisely those with critical angles $3\lambda_S - 5\lambda_U - 7\lambda_N + 7\varpi_6^* + (2-q)\varpi_7^* + q\varpi_8^*$. Curiously, the present position of Uranus corresponds to the only place in the 19.218–19.22 AU interval where the resonant components do not quite overlap. This explains the very strong dependence of the Lyapunov time on the initial conditions and integration approximations, found by Sussman and Wisdom (1988).

The uncertainty in the current semimajor axis of Uranus is 6×10^{-6} AU and the physical size of the planet is $\sim 10^{-4}$ AU. It is therefore impossible to conclude what is the real Lyapunov time of Uranus and what is the multiplet resonant component that really affects its motion. However, the results by Murray and Holman leave little room for doubt about the importance of three-body resonances in the giant planets' motion. With the dynamical model that will be explained in Chapter 12, Murray and Holman estimated that, because of its chaotic motion, Uranus could increase its eccentricity up to a Saturn-crossing value, but in a time not shorter than 10^{18} y, that is 8 orders of magnitude longer than the physical lifetime of the Sun. Therefore, the system of the giant planets is *practically stable.*

Chapter 11

SECULAR DYNAMICS INSIDE MEAN MOTION RESONANCES

11.1 Successive elimination of harmonics

In Chapter 9 we have seen that the mean motion resonances of low order with a major planet have, to a first approximation, a dynamical structure typical of a modulated pendulum: they are characterized by the presence of a central region of "regular" motion, surrounded by a chaotic layer. The goal of this chapter is to study the dynamics inside this central region on timescales that are much longer than the libration period inside the mean motion resonance. For this purpose one must fully take into account also the secular part of the perturbation (9.17). This leads us to study resonances inside a resonance (or resonances inside a resonance inside a resonance as in the case of Pluto), namely highly nonlinear dynamics.

As discussed in Chapter 4, the elimination of the angular dependence of a resonant Hamiltonian cannot be achieved using the Lie algorithm, even in the integrable case. Instead, it can be done with the introduction of Arnold action–angle variables. In these variables, an integrable resonant Hamiltonian is written as a function of the sole action variables and can therefore be used as the starting basis for further investigation of the dynamics. Therefore, the study of a problem characterized by a hierarchical interaction of several resonances must require the introduction of a sequence of transformations to suitable Arnold action–angle variables. For these problems, the use of Arnold variables allows us to revisit a perturbation approach originally due to Delaunay (1867), based on the successive elimination of perturbation harmonics up

to a given order.

The principle of the algorithm is simple. Consider a Hamiltonian $\mathcal{H}(\mathbf{I}, \boldsymbol{\varphi})$ and expand it in a Fourier series

$$\mathcal{H} = \mathcal{H}_0(\mathbf{I}) + \sum_{\mathbf{m}} h_{\mathbf{m}}(\mathbf{I}) \exp(\iota \mathbf{m} \cdot \boldsymbol{\varphi}) \ . \tag{11.1}$$

Then choose one harmonic, say with index $\bar{\mathbf{m}}$, and its multiples, and consider the partial Hamiltonian

$$\mathcal{H}_{\bar{\mathbf{m}}} = \mathcal{H}_0(\mathbf{I}) + \sum_{k \in \mathbf{Z}} h_{k\bar{\mathbf{m}}}(\mathbf{I}) \exp(\iota k \bar{\mathbf{m}} \cdot \boldsymbol{\varphi}) \ . \tag{11.2}$$

Since this Hamiltonian is integrable, introduce Arnold action–angle variables $\mathbf{J}, \boldsymbol{\psi}$ in order to write it as a function only of the new actions, i.e. $\mathcal{H}_0'(\mathbf{J})$. The harmonics considered in (11.2) are completely eliminated. Afterwards, write the remainder

$$\mathcal{R} = \sum_{\mathbf{m} \neq k\bar{\mathbf{m}}, k \in \mathbf{Z}} h_{\mathbf{m}}(\mathbf{I}) \exp(\iota \mathbf{m} \cdot \boldsymbol{\varphi}) \tag{11.3}$$

in the new variables $\mathbf{J}, \boldsymbol{\psi}$ and expand it in a Fourier series of the new angles $\boldsymbol{\psi}$:

$$\mathcal{R} = \sum_{\mathbf{m} \neq k\bar{\mathbf{m}}, k \in \mathbf{Z}} h_{\mathbf{m}}'(\mathbf{J}) \exp(\iota \mathbf{m} \cdot \boldsymbol{\psi}) \ . \tag{11.4}$$

The new Hamiltonian $\mathcal{H}_0'(\mathbf{J}) + \mathcal{R}(\mathbf{J}, \boldsymbol{\psi})$ is formally equivalent to (11.1); thus the algorithm can be iterated, as sketched in Fig. 11.1, in order to eliminate all the principal harmonics.

Although apparently simple, the algorithm sketched above hides several conceptual difficulties. The first one concerns the sequential order by which the harmonics should be eliminated. The experience suggests that for an efficient and effective practical implementation one should eliminate at each iteration the harmonic that has the most important influence on the motion of the action variables; this is in general the harmonic with the largest coefficient among the resonant ones, which is usually also that associated with the shortest librational timescale (as the latter is proportional to the square root of the harmonic's coefficient; see Section 4.1). When two harmonics are resonant in different parts of the action space, it is irrelevant which of the two is first eliminated. We will see in the next section that, in applying the algorithm to mean motion resonance dynamics, the suitable order of successive elimination of harmonics imposes itself in a quite natural way.

The second conceptual difficulty is that, when a resonant harmonic is eliminated, the introduction of Arnold action–angle variables is not global over the

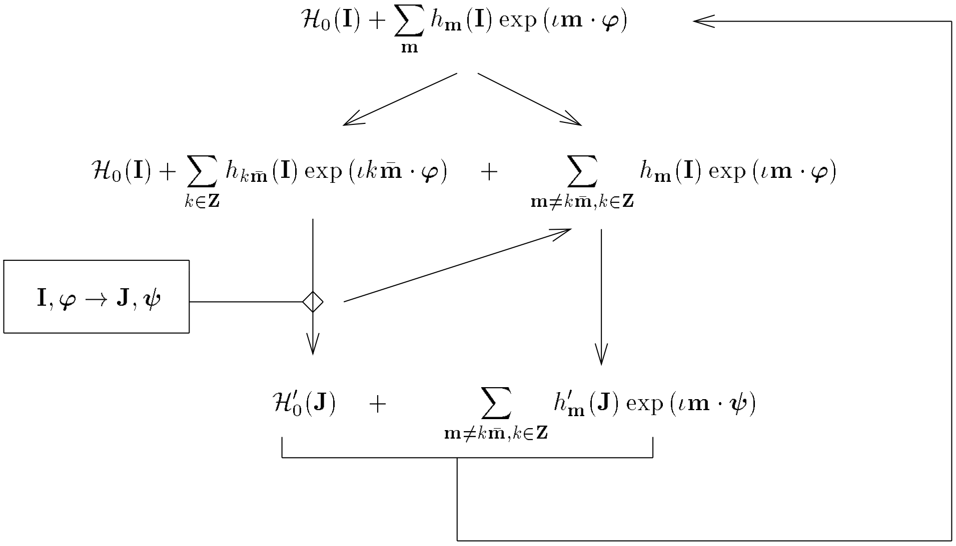


Figure 11.1: Sketch of the algorithm of successive elimination of harmonics by the introduction of Arnold action–angle variables. See text.

entire phase space (see Section 4.2). In fact, the algorithm imposes a fragmentation of the phase space into disconnected regions, according to the location of the separatrices of the considered resonance. Different action–angle variables are introduced in the regions of libration and circulation of the resonant angle, leading to the definition of different new Hamiltonians $\mathcal{H}_0'(\mathbf{J}) + \mathcal{R}(\mathbf{J}, \boldsymbol{\psi})$, one for each region. This corresponds to the real fact that the dynamics inside a resonance may be completely different from the dynamics just outside of it. Therefore, the new Hamiltonians are defined on domains of the actions $\mathbf{J}$ that are delimited by well-defined boundaries, which are dictated by the separatrices of the previously eliminated resonance. When a new harmonic is considered, the actions $\mathbf{J}$ are no longer constant, so that some trajectories hit the boundaries of the action domain. Evidently, for these trajectories it is not possible to introduce new Arnold action–angle variables so that the algorithm of successive elimination of harmonics cannot be iterated. As a consequence, the algorithm does not allow the study of regions of phase space close to the separatrices of the resonances. However, this is not a fault of the algorithm, because these portions of phase space are expected to correspond to chaotic regions (which, as we know from Chapter 4, originate around the separatrices of an integrable resonant approximation when a perturbation is introduced).

A third conceptual difficulty is that in general the already eliminated harmonics reappear in the course of the algorithm, when other harmonics are eliminated. However, under the assumption that the original Hamiltonian is analytic and has a quasi-integrable form (see 2.1), it is easy to show that the eliminated harmonics can reappear only with coefficients that are much smaller than the original ones. In this sense, the method of successive elimination of the harmonics can be considered as a genuine perturbation theory. Morbidelli and Giorgilli (1993) proved that, if the perturbation parameter is small enough, the algorithm converges on a set of KAM tori.

The last difficulty is related to the technical implementation of the algorithm. As remarked in Section 4.2, it is very hard, if not impossible, to write the Hamiltonian $\mathcal{H}_0'(\mathbf{J})$ in explicit form. Conversely, on each torus $\mathbf{J}$ = constant it is possible to compute the Hamiltonian, its derivatives, as well as the remainder $\mathcal{R}(\mathbf{J}, \psi)$ and its Fourier expansion, with a seminumerical procedure. Therefore, the algorithm of successive elimination of harmonics can be implemented by evaluating, at each iteration, the new Hamiltonian over a grid in the action space. The Hamiltonian flow is then computed in every required point of the phase space, using an interpolation algorithm. This practical implementation technique is explained in detail in Morbidelli (1993b), in the framework of a nontrivial example.

With reference to the algorithm of successive elimination of harmonics, we can now investigate in a structured approach the complex dynamics existing inside low-order mean motion resonances.

11.2 The mean motion resonant dynamical system

For a $k_{\bar{\jmath}}\dot{\lambda}_{\bar{\jmath}} - k\dot{\lambda} \sim 0$ mean motion resonance between an asteroid and the $\bar{\jmath}$-th planet (with $k_{\bar{\jmath}} \neq k$), we start from the Hamiltonian (9.15). Remember that (9.15) is the nonsimplified mean motion resonant normal form, where the secular motion of the perturbing planets is also taken into account. As a first step of the process of successive elimination of harmonics we need to find an integrable approximation of the dynamics. As discussed in Section 9.4, if the inclination is not excessively large and the eccentricity of the asteroid is significantly larger than those of the planets, the resonant harmonic with angle

$$(k_{\bar{\jmath}} - k)\sigma = k_{\bar{\jmath}}\lambda_{\bar{\jmath}} - k\lambda + (k_{\bar{\jmath}} - k)p$$

dominates in width over all other harmonics in the resonant multiplet. Therefore, we define an integrable approximation $\mathcal{H}_\sigma$ by retaining from the perturbation $\mathcal{H}_1$ in (9.15) only this harmonic and its multiples. This integrable

approximation has the form:

$$\mathcal{H}_\sigma = \mathcal{H}_0(\tilde{\Lambda}_{\bar{j}}, N, S, S_z) + \sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j}) + \varepsilon \mathcal{F}_\sigma(S, (k_{\bar{j}} - k)\sigma, N, S_z) \ , \quad (11.5)$$

where $\mathcal{H}_0$ is given by (9.7) and $\mathcal{F}_\sigma$ retains all the terms in the perturbation that do not depend on angles other than σ.

A complication is that the Fourier expansion of $\mathcal{H}_1$ in (9.15) does not converge if the asteroid's eccentricity is large enough to allow collisions with the resonant planet for some values of σ (see Section 9.1.1). In these cases, the function $\mathcal{F}_\sigma$ can still be computed for every value of σ that does not allow collisions, by writing $\mathcal{H}_1$ as a function of the variables (9.5) without using series expansions, and by averaging it over $\sigma_z, \boldsymbol{\varpi}^*$ and $\boldsymbol{\Omega}^*$. The local evaluation of $\mathcal{H}_1$ and of its derivatives, avoiding series expansions, can be obtained following Ferraz-Mello and Sato (1989) and Moons (1993, 1994).

The Hamiltonian (11.5) is formally equivalent to that of the planar circular restricted problem, but in (9.4) S_z was assumed equal to 0, while now $\mathcal{F}_\sigma$ depends parametrically on the value of S_z (constant of motion for 11.5). Recall that S_z depends essentially on the value of the inclination.

Because the Hamiltonian (11.5) is integrable, we introduce Arnold action–angle variables. According to Section 4.2 the transformation has the form

$$\begin{aligned} \psi_\sigma &= \frac{2\pi}{T_\sigma} t \ , & J_\sigma &= \frac{1}{2\pi} \oint S \, \mathrm{d}\sigma \\ \psi_z &= \sigma_z - \varrho_z(\psi_\sigma, J_\sigma, J_z, J_\nu) \ , & J_z &= S_z \\ \psi_\nu &= \nu - \varrho_\nu(\psi_\sigma, J_\sigma, J_z, J_\nu) \ , & J_\nu &= N \ , \end{aligned} \quad (11.6)$$

where the first line indicates, with an abridged notation, that the relationship between the original (S, σ) and the new (J_σ, ψ_σ) action–angle variables is that detailed in Section 4.2, and the functions ϱ_z and ϱ_ν are periodic with zero average in ψ_σ.

Because we are interested in the dynamics *inside* the mean motion resonance, we compute the separatrices of $\mathcal{H}_\sigma$ in (σ, S, S_z, N) space, and introduce the variables (11.6) only for trajectories enclosed by the separatrices, so that $|T_\sigma|$ is the period of libration and J_σ is the normalized area bounded by the librating trajectory in the (S, σ) plane.[1]

[1]As sketched in Fig. 9.3, in resonances of first order in the eccentricity σ might circulate on some resonant trajectories, but action–angle variables can still be introduced by (11.6), with $|T_\sigma|$ denoting the period of the cycle in (S, σ). Using the convention introduced in Section 4.2, T_σ is positive for exterior resonances (because the libration in the (σ, S) plane is clockwise) while it is negative for interior resonances.

When σ librates around some value σ_0, the dependence of σ on ψ_σ is of type $\sigma = \sigma_0 + \varrho(\psi_\sigma)$, with ϱ periodic with zero average. Thus, it follows from (9.5) that $\sigma_z = \sigma_0 + \varrho(\psi_\sigma) - p + q = \sigma_0 + \varrho(\psi_\sigma) + \omega$ and $\nu = -\sigma_0 - \varrho(\psi_\sigma) + p = -\sigma_0 - \varrho(\psi_\sigma) - \varpi$. Recalling that the functions ϱ_z and ϱ_ν in (11.6) are also periodic with zero average in ψ_σ, this implies that ψ_z and $-\psi_\nu$ can be identified respectively with the argument of perihelion and the longitude of perihelion of the asteroid, averaged over the libration period in the resonance and shifted by σ_0.

As a function of (11.6), $\mathcal{H}_\sigma$ depends only on the action variables, i.e.

$$\mathcal{H}_\sigma \equiv \sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j}) + \mathcal{F}_0(J_\sigma, J_z, J_\nu) \ . \tag{11.7}$$

The variables (11.6) are also introduced in the part of the original perturbation that was not included in (11.5), namely in

$$\mathcal{H}_1 - \mathcal{F}_\sigma = \mathcal{F}_1(\psi_\sigma, \psi_z, \psi_\nu, \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*, J_\sigma, J_z, J_\nu) \ . \tag{11.8}$$

As (11.7) depends only on the action variables, the frequencies:

$$\dot{\psi}_\sigma = \frac{\partial \mathcal{F}_0}{\partial J_\sigma} \ , \quad \dot{\psi}_z = \frac{\partial \mathcal{F}_0}{\partial J_z} \ , \quad \dot{\psi}_\nu = \frac{\partial \mathcal{F}_0}{\partial J_\nu} \tag{11.9}$$

are constant and depend only on (J_σ, J_z, J_ν). We can therefore look at the different possible resonances in order to decide which harmonic of $\mathcal{F}_1$ should be considered in the next step of the study.

Notice that the location of the resonances, once determined in (J_σ, J_z, J_ν) space, can be easily imaged in (a, e, i) space through a section at $\sigma = \sigma_0$. The procedure is equivalent to that used in Chapter 9 for plotting the separatrices of the mean motion resonance on the (a, e) plane. In fact, the values of the three actions J_σ, S_z, N identify a cycle on the (S, σ) plane, along which S assumes two values S_1 and S_2 corresponding to $\sigma = \sigma_0$. The sets of values (S_1, S_z, N) and (S_2, S_z, N) are then respectively transformed into the sets (a_1, e_1, i_1) and (a_2, e_2, i_2) – one on each side of the resonance center – using (9.5) and the definition (1.69) of the modified Delaunay variables.

11.2.1 Secondary resonances

The terminology *secondary resonance* was introduced by Lemaître and Henrard (1990) to denote the resonance between the frequency of ψ_σ and the frequency of a combination of the secular angles, whose corresponding harmonic appears in (11.8).

In a problem where the inclination of the asteroid is taken into account, there is one family of secondary resonances whose corresponding harmonics satisfy the second D'Alembert rule (see Section 1.9.3) and have coefficients independent of the planetary eccentricities and inclinations. This family corresponds to the commensurabilities

$$m\dot{\psi}_\sigma + 2n\dot{\psi}_z = 0 \ , \tag{11.10}$$

with integer m, n. As $\dot{\psi}_z$ is the average frequency of the argument of perihelion ω, this family corresponds to the resonances between the libration period of σ and the circulation period of ω.

Taking into account also harmonics whose coefficients are powers of the planetary eccentricities or inclinations, many other secondary resonances appear. Among them, we distinguish the family which corresponds to the commensurabilities

$$m\dot{\psi}_\sigma + n(\dot{\psi}_\nu + g_j) = 0 \ , \tag{11.11}$$

and that associated with the relations

$$m\dot{\psi}_\sigma + n(\dot{\psi}_z + \dot{\psi}_\nu + s_j) = 0 \ . \tag{11.12}$$

Because $\dot{\psi}_\nu$ is equal to the average frequency of $-\varpi$, the first family corresponds to commensurabilities between the libration period in the mean motion resonance and the circulation period of $\varpi - \varpi_j^*$, while the second family corresponds to commensurabilities between the libration period in the mean motion resonance and the circulation period of $\Omega - \Omega_j^*$. Other families of secondary resonances involve combinations of perihelia and nodes.

By definition, in secondary resonances the secular angles must circulate with a period that is of the order of the libration period in the mean motion resonance; therefore, the secondary resonances are typically located in regions were there are no low-order secular resonances, and they therefore dominate the local dynamics. These regions are in general characterized by values of the eccentricity that are low to moderate. As an example, Fig. 11.2 shows the upper border (light dotted curve) of the region concerned by the resonances of the family (11.11) with $m = 1$ and $n \leq 5$ in the 3/1 resonance with Jupiter: these resonances are all located at $e < 0.1$. See also Figs. 11.10, 11.17 and 11.18 for the location of secondary resonances in the 2/1 and 3/2 resonance with Jupiter and in the 2/3 resonance with Neptune, respectively.

As one sees from the above quoted figures, the secondary resonances typically cluster in a narrow region. As a consequence, it does not make much sense to try to successively eliminate their corresponding harmonics or to develop single resonance models to study their dynamics. It is more efficient

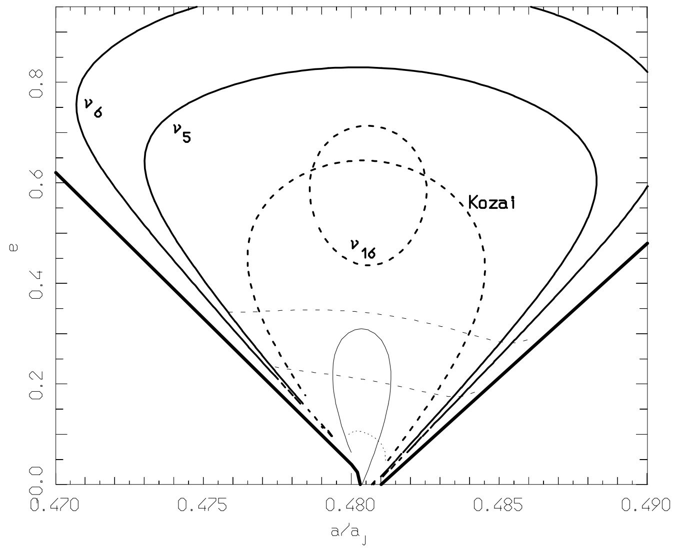


Figure 11.2: The location of the main resonances inside the 3/1 commensurability. The two thick lines on each side of the plot denote the separatrices of the 3/1 resonance, plotted for $\sigma = 90$ degrees. Inside the 3/1 resonance, the bold solid curves show the location of the ν_5 and ν_6 secular resonances; the bold dashed curves refer to the ν_{16} and to the Kozai secular resonances; the light solid curve marks the location of the secular resonance $\dot{\vartheta}_\nu + g_6 + 2\dot{\vartheta}_z = 0$; the light dotted curve refers to the secondary resonance $\dot{\psi}_\sigma + 5(\dot{\psi}_\nu + g_6) = 0$, which bounds the region concerned by secondary resonances of type (11.11) of order not larger than 5; the light dashed curves denote the location of the three-body resonances $2\dot{\psi}_\sigma + 3\dot{\varphi}_{\rm JS} + 9\dot{\psi}_\nu = 0$ and $3\dot{\psi}_\sigma + 3\dot{\varphi}_{\rm JS} + 9\dot{\psi}_\nu = 0$, respectively at larger and smaller eccentricity. Remember that $\dot{\psi}_\sigma < 0$ for interior resonances.

to select a family of secondary resonances – say for simplicity that given by (11.11) with a given j – to step back to the original Hamiltonian (9.15) written in the variables (9.5), and to retain all the harmonics that depend on σ and/or on $\nu + \varpi_j^*$. In this way one obtains a two-degree of freedom system of type

$$\mathcal{H}_{\sigma,\nu} = \mathcal{H}_0(\tilde{\Lambda}_{\bar{j}}, N, S, S_z) + g_j \Lambda_{g_j} + \varepsilon \mathcal{F}_{\sigma,\nu}(S, (k_{\bar{j}} - k)\sigma, N, \nu + \varpi_j^*, S_z) \ , \quad (11.13)$$

which can be studied using Poincaré sections. As an example, Fig. 11.3 shows the Poincaré sections on six energy levels of (11.13) for the 2/1 resonance with Jupiter, under the assumption that Jupiter is on a fixed eccentric orbit (set $g_j = 0$ in the expression of $\mathcal{H}_{\sigma,\nu}$). As one sees, the individual secondary resonances overlap, creating a chaotic layer that extends approximately over an

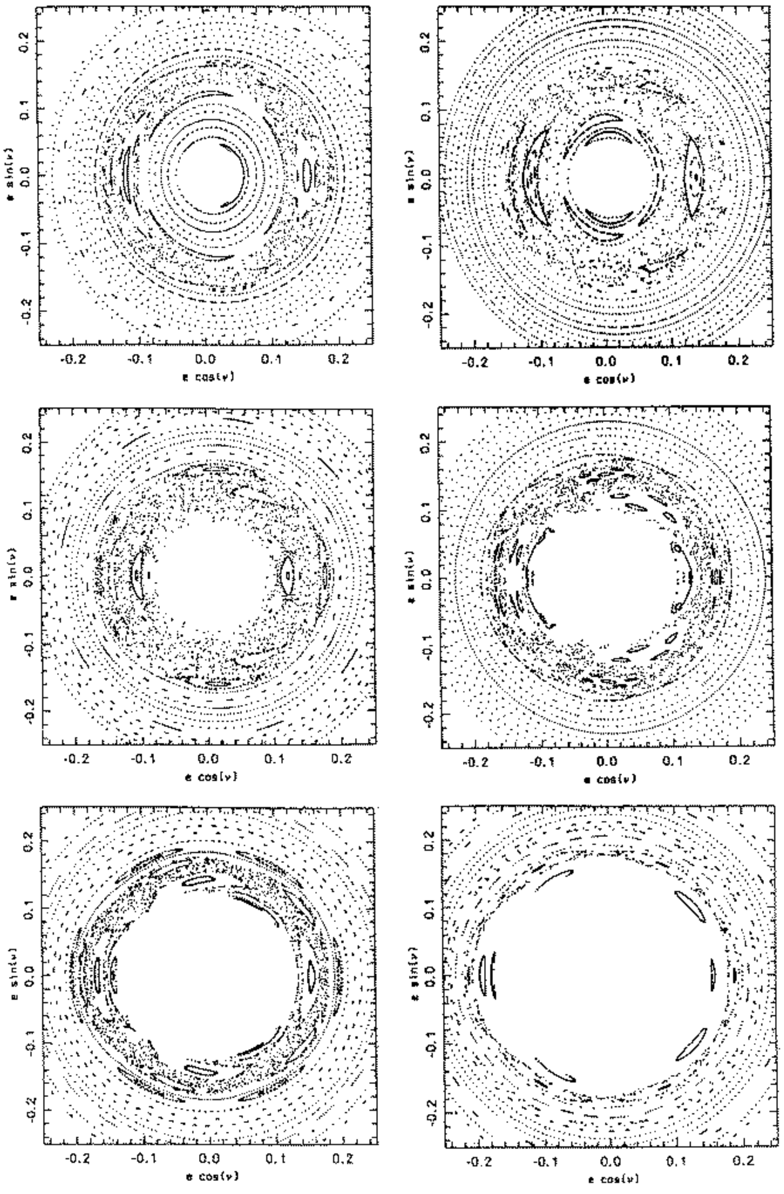


Figure 11.3: Poincaré sections of (11.13) at $\sigma = 0, \dot{\sigma} < 0$. Jupiter is assumed to be in a fixed elliptic orbit with eccentricity equal to 0.0485. On each panel, the amplitude of libration in the 2/1 resonance is zero for the most internal orbit, and increases with increasing eccentricity. Reprinted from Fig. 6 of Moons and Morbidelli (1993), with permission from Kluwer Academic Publishers.

interval 0.10–0.15 in eccentricity. At larger eccentricity the motion is essentially regular since no low-order secondary resonances exist (see Fig. 11.10). An equivalent study can be done for each family of secondary resonances, thus achieving a view of the extension of the chaotic regions generated by these resonances.

11.2.2 The Kozai dynamics

In the regions where there are no secondary resonances of low order, the dynamics is dominated by the *secular* harmonics, namely by the harmonics of $\mathcal{F}_1$ that do not depend on ψ_σ. Neglecting the terms dependent on ψ_σ in (11.8), the Hamiltonian that describes the secular dynamics inside the mean motion resonance has the form

$$\mathcal{H}_{\rm SEC} = \sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j}) + \mathcal{F}_0(J_\sigma, J_z, J_\nu) + \bar{\mathcal{F}}_1(\psi_z, \psi_\nu, \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*, J_\sigma, J_z, J_\nu) \tag{11.14}$$

where $\bar{\mathcal{F}}_1$ is the average of $\mathcal{F}_1$ over ψ_σ. The action J_σ is a constant of motion for (11.14). Figure 11.4 shows, for the 3/1 resonance with Jupiter, the sets of values (a, e) that correspond to $J_\sigma =$constant when $i = 0$ and $\sigma = 90°$ (the libration center's value). The pairs of dotted curves, symmetrically placed with respect to the center of the resonance, correspond to the same value of J_σ. These curves do not change significantly for different values of i. As one sees, the preservation of J_σ during the secular evolution of the eccentricity implies that the amplitude of oscillation of the semimajor axis (but *not* of σ!) remains almost constant. In the above quoted figures, increasing values of J_σ correspond to pairs of dotted curves with increasing distance from the resonance center. In the limit $J_\sigma = 0$ the corresponding curves collapse onto the curve denoting the center of the resonance; in fact, $J_\sigma = 0$ corresponds to orbits that sit on the stable equilibrium point of the mean motion resonance. If one restricts the study of the secular dynamics to $J_\sigma = 0$ (see for instance Yoshikawa, 1990, 1991), the transformation (11.6) for ψ_z and ψ_ν becomes the identity and the Hamiltonian (11.14) is simply obtained from the original (9.15) by setting $\sigma = \sigma_0$ and $\Lambda = \Lambda^*(S_z, N)$, where $\Lambda^*(S_z, N)$ denotes the value of Λ at the stable equilibrium point of (11.5), which weakly depends on S_z and N.

For whatever value of J_σ, the Hamiltonian (11.14) is not integrable as it depends on several angles. The study of its dynamics requires again the hierarchical consideration of its main harmonic terms. For the D'Alembert rules, all the harmonics that depend on $\psi_\nu, \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*$ must have coefficients that are powers of the planetary eccentricities or inclinations, which are in

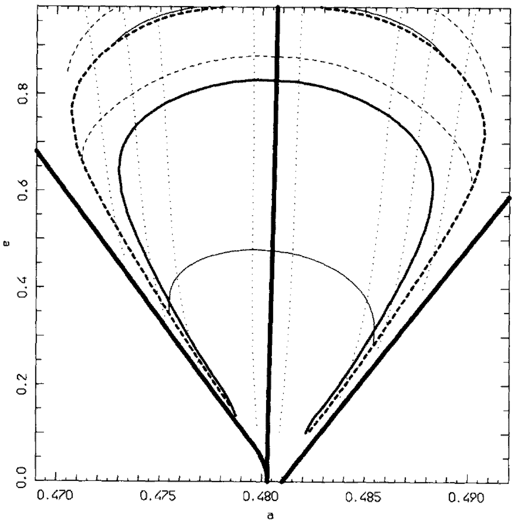


Figure 11.4: Secular dynamics in the 3/1 resonance with Jupiter in the (a, e) plane at $i = 0°$. The pair of dotted lines, symmetrically placed with respect to the resonance center, denote the sets of (a, e) values that correspond to $J_\sigma =$ constant at $\sigma = 90°$. The bold solid and dashed curves denote respectively the location of the ν_5 and of the ν_6 resonances. The pair of light solid curves and light dashed curves mark respectively the separatrices of the two secular resonances. The unit on the abscissa is Jupiter's semimajor axis. Reprinted from Fig. 6 of Moons and Morbidelli (1995), with permission from Academic Press.

general smaller than those of the asteroid. Thus, as for the secular problem outside mean motion resonances (see Chapter 8), the leading harmonic part of the Hamiltonian is the one depending solely on the argument of perihelion, namely on ψ_z in the variables (11.6). In fact these harmonics are the only ones with nonzero coefficients in the case where the eccentricities and inclinations of the planets are all null.

Therefore, the next step in the process of elimination of harmonics is to eliminate the terms in ψ_z. For this purpose, we first write

$$\mathcal{H}_{\rm SEC} = \sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j}) + \mathcal{F}_0 + \mathcal{F}_z + (\bar{\mathcal{F}}_1 - \mathcal{F}_z) \ , \tag{11.15}$$

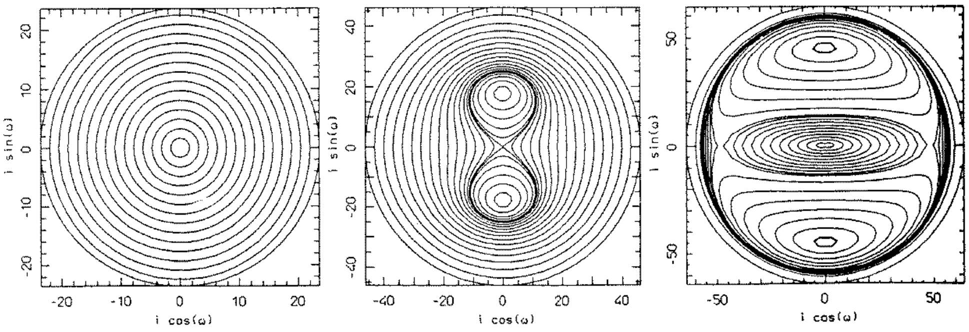


Figure 11.5: Typical portraits of the dynamics of (11.16) for three values of the constant action J_ν and setting $J_\sigma = 0$. The coordinates J_z, ψ_z have been converted to inclination i and argument of perihelion ω. Reprinted from Fig. 5 of Morbidelli and Moons (1993), with permission from Academic Press.

where $\mathcal{F}_z$ is the average of $\bar{\mathcal{F}}_1$ over $\boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*$;[2] then we consider the Hamiltonian

$$\mathcal{H}_z = \sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j}) + \mathcal{F}_0(J_\sigma, J_z, J_\nu) + \mathcal{F}_z(\psi_z, J_\sigma, J_z, J_\nu) \ , \qquad (11.16)$$

which is integrable because it depends on only one angle (ψ_z). The Hamiltonian (11.16) is the equivalent of the Kozai Hamiltonian (8.16), but for the secular dynamics inside a mean motion resonance.

The dynamics of (11.16) preserves the value of the action $J_\nu \equiv N$. On a surface in (a, e, i) space defined by $J_\sigma =$ constant[3] and $\sigma = \sigma_0$, the condition $J_\nu =$ constant gives a relationship between the eccentricity and the inclination, which is represented by the dotted curves in Fig. 11.6 for the 2/1 mean motion resonances with Jupiter. The eccentricity and the inclination therefore exhibit coupled oscillations, correlated with the motion of J_z, ψ_z. The evolution of J_z, ψ_z can be easily represented plotting the level curves of (11.16), J_σ and J_ν playing the role of parameters. As an example, Fig. 11.5 shows these level curves for the 2/1 resonance with Jupiter, in the case $J_\sigma = 0$ and for three values of J_ν. For J_ν sufficiently large, one sees a typical resonant portrait; this shows that the equivalent of the Kozai resonance is present inside the 2/1 mean motion resonance with Jupiter. This property is common to

[2] The angle ψ_ν disappears with the average over $\boldsymbol{\varpi}^*$ and $\boldsymbol{\Omega}^*$ because of the D'Alembert rules. Neglecting terms of order $(e_j, i_j)^2$, the average over $\boldsymbol{\varpi}^*$ and $\boldsymbol{\Omega}^*$ is simply obtained by setting $e_j = i_j = 0$ in $\bar{\mathcal{F}}_1$.

[3] Which can often be approximated by a surface $a =$ constant, in order to make the practical computations easier.

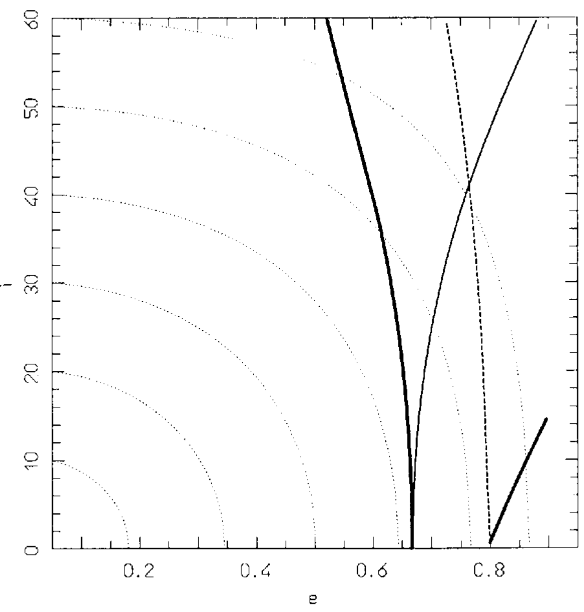


Figure 11.6: Secular dynamics at the center of the 2/1 resonance ($J_\sigma = 0$). The dotted curves denote the sets of (e, i) values that correspond to $J_\nu =$constant. The light solid curve represents the stable family of equilibrium points of the argument of perihelion ($\psi_z = 90°, 270°$); the dashed curve represents the unstable family ($\psi_z = 0°, 180°$). The two bold curves denote the separatrices of the Kozai resonance, computed at $\psi_z = 90°$. Reprinted from Fig. 6 of Morbidelli and Moons (1993), with permission from Academic Press.

many other mean motion resonances (see Jefferys and Standish, 1972). Note that the dynamics is π-periodic in ψ_z, as a consequence of the fact that, for the D'Alembert rules, the Hamiltonian (11.16) can have only harmonics $\exp[\iota m \psi_z]$ with even m. Denote by ψ_z^0 the value of ψ_z for one of the stable equilibrium points; the separatrix of the Kozai resonance intersects the axis $\psi_z = \psi_z^0$ in two values of J_z, say $J_z^{(1)}, J_z^{(2)}$, which depend on J_ν. Remembering that $J_z \equiv S_z$ and $J_\nu \equiv N$, one can convert the values $(J_z^{(1)}(J_\nu), J_\nu)$ and $(J_z^{(2)}(J_\nu), J_\nu)$ into $(e_1(N), i_1(N))$ and $(e_2(N), i_2(N))$ on the surface corresponding to the considered value of J_σ, thus obtaining the image of the separatrices of the Kozai resonance on an eccentricity vs. inclination plot. The equivalent procedure can be done also for the families of stable and unstable equilibrium points of (11.16). Figure 11.6 gives an example of this representation, again for the 2/1

resonance.

In each of the dynamical regions of (11.16) (positive or negative circulation, libration of ψ_z), we can now proceed to introduce action-angle variables

$$\begin{aligned} \vartheta_z &= \frac{2\pi}{T_z}t\ , & \Theta_z &= \frac{1}{2\pi}\oint J_z\, \mathrm{d}\psi_z \\ \vartheta_\sigma &= \psi_\sigma - \varrho'_\sigma(\vartheta_z, \Theta_\sigma, \Theta_z, \Theta_\nu)\ , & \Theta_\sigma &= J_\sigma \\ \vartheta_\nu &= \psi_\nu - \varrho'_\nu(\vartheta_z, \Theta_\sigma, \Theta_z, \Theta_\nu)\ , & \Theta_\nu &= J_\nu\ , \end{aligned} \tag{11.17}$$

(where the same notation as formula 11.6 has been used) in order to make $\mathcal{H}_z$ dependent only on the action variables i.e.

$$\mathcal{H}_z \equiv \sum_j (g_j \Lambda_{g_j} + s_j \Lambda_{s_j}) + \mathcal{K}_0(\Theta_\sigma, \Theta_z, \Theta_\nu)\ . \tag{11.18}$$

When the new variables (11.17) are introduced into the neglected term $\mathcal{K}_1 \equiv \bar{\mathcal{F}}_1 - \mathcal{F}_z$, the complete secular Hamiltonian (11.15) takes the form:

$$\mathcal{H}_{\mathrm{SEC}} = \mathcal{K}_0(\Lambda_{g_j}, \Lambda_{s_j}, \Theta_\sigma, \Theta_z, \Theta_\nu) + \mathcal{K}_1(\vartheta_z, \vartheta_\nu, \boldsymbol{\varpi}^*, \boldsymbol{\Omega}^*, \Theta_\sigma, \Theta_z, \Theta_\nu)\ . \tag{11.19}$$

The “unperturbed” frequencies of the Hamiltonian (11.19) are now

$$\dot{\vartheta}_\sigma = \frac{\partial \mathcal{K}_0}{\partial \Theta_\sigma}\ , \quad \dot{\vartheta}_z = \frac{\partial \mathcal{K}_0}{\partial \Theta_z}\ , \quad \dot{\vartheta}_\nu = \frac{\partial \mathcal{K}_0}{\partial \Theta_\nu}\ , \tag{11.20}$$

together with the planetary secular frequencies g_j and s_j. These frequencies allow one to localize the resonances associated to the harmonics that appear in $\mathcal{K}_1$. Among these resonances we distinguish two families:

(i) *The perihelion secular resonances*: these resonances correspond to the relationships $m\dot{\vartheta}_\nu + \sum_j m_j g_j \sim 0$, where $m = \sum_j m_j$ for the D'Alembert rules; in fact it is enough to notice that the transformation between ϑ_ν and ψ_ν in (11.17) is close to the identity, and to remember that $-\psi_\nu$ represents the average over the libration period of σ of the asteroid's longitude of perihelion ϖ. We detail in Section 11.2.3 the secular resonances $\dot{\vartheta}_\nu + g_j \sim 0$, for the main frequencies g_5 and g_6.

(ii) *The nodal secular resonances*: when ψ_z circulates, the transformation between ϑ_z and ψ_z in (11.17) is close to the identity; remembering that ψ_z is the mean argument of perihelion over the σ-libration, the nodal secular resonances are given by the relations $m(\dot{\vartheta}_z + \dot{\vartheta}_\nu) + \sum_j m_j s_j \sim 0$, where again $m = \sum_j m_j$. In Section 11.2.4 we consider the resonance $\dot{\vartheta}_z + \dot{\vartheta}_\nu + s_6 \sim 0$ for the mean motion resonances in the asteroid belt and $\dot{\vartheta}_z + \dot{\vartheta}_\nu + s_8 \sim 0$ for the Kuiper

belt. In the region where ψ_z librates, the nodal secular resonances become those given by $m\dot{\vartheta}_\nu + \sum_j m_j s_j \sim 0$.

However, many other secular resonances may occur; for example Pluto, which is in the Kozai resonance inside the 2/3 mean motion resonance with Neptune, is also in the resonance between the libration period of ψ_z and the circulation period of $\vartheta_\nu + \Omega_8^*$ (Milani *et al.*, 1989).

When the location of the resonances in $(\Theta_\sigma, \Theta_z, \Theta_\nu)$ space is determined, one can compute its image in (a, e, i) space, with the following procedure: one first computes the value of J_z that corresponds to Θ_z when $\psi_z = \psi_z^0$, and then tranforms the resulting values of $(J_z, J_\nu \equiv \Theta_\nu, J_\sigma \equiv \Theta_\sigma)$ into (a, e, i) as explained above for the separatrices of the Kozai resonance.

Note that, if one restricts to the planar problem, all the harmonics of ψ_z in (11.16) have zero coefficients, as the latter are at least quadratic in the asteroid's inclination. Thus, the transformation (11.17) becomes the identity on all the variables, and the Hamiltonian (11.19) is the same as (11.14). This allows the simplification of the computational procedure, but on the other hand severely limits the validity of the results to the plane $i = 0$. For a nonzero inclination, Yoshikawa (1990, 1991) has simplified the procedure explained in this section by averaging (11.14) over ψ_z and approximating the trasformation (11.17) with the identity; in this way he obtained a Hamiltonian that is formally equivalent to (11.19), but that is accurate only in the region where ψ_z rapidly circulates and induces negligible oscillations of S_z. Because the Kozai resonance hits the plane $i = 0$ at large eccentricity (see Figs. 11.6, 11.9, 11.19), his study is therefore valid only for small to moderate values of e.

11.2.3 Perihelion secular resonances

Figures 11.10, 11.2 and 11.17 show the location for $i = 0$ of the resonances $\dot{\vartheta}_\nu + g_5 = 0$ and $\dot{\vartheta}_\nu + g_6 = 0$ (called also respectively ν_5 and ν_6) in the 2/1, 3/1 and 3/2 mean motion resonances in the asteroid belt. Recall from Chapter 8 that outside the mean motion resonances these secular resonances occur only for large inclination (with the exception of the ν_6 resonance at $a \sim 2.0$–2.1 AU). *The secular dynamics is therefore completely different inside and outside mean motion resonances.* This feature of degenerate dynamics is well understood using the approach of successive elimination of harmonics: in fact, as discussed in Section 11.1, the Arnold action–angle variables that are suitable for the regions of circulation and libration of σ are completely different, as well as the resulting secular Hamiltonians.

Analogously, Fig. 11.18 shows the location of the resonance $\dot{\vartheta}_\nu + g_8 = 0$ (also called ν_8) in the 2/3 resonance with Neptune in the Kuiper belt.

To study a resonance of type $\dot{\vartheta}_\nu + g_j \sim 0$ we isolate in (11.19) the harmonic of $\mathcal{K}_1$ with argument $\sigma_j = \vartheta_\nu + \varpi_j^*$, and consider the integrable model

$$\mathcal{H}_j = g_j \Lambda_{g_j} + \mathcal{K}_0(\Theta_\sigma, \Theta_z, \Theta_\nu) + \mathcal{K}_1^j(\Theta_\sigma, \Theta_z, \Theta_\nu, \sigma_j) \ . \tag{11.21}$$

The function $\mathcal{K}_1^j$ is the average of $\mathcal{K}_1$ over all the angles except $\vartheta_\nu + \varpi_j^*$. The dynamics of (11.21) preserves the value of Θ_σ and Θ_z, while only $\Theta_\nu \equiv N$ evolves with the circulation/libration of σ_j. The motion of N, σ_j is represented, as usual, by the level curves of (11.21), Θ_σ and Θ_z playing the role of parameters. As an example, Fig. 11.7 shows the dynamics of the resonance $\dot{\vartheta}_\nu + g_5 = 0$ for four values of Θ_σ and $\Theta_z = 0$ in the 3/1 mean motion resonance with Jupiter. As one sees, the phase portrait shows a huge libration island that dominates almost the entire (N, σ_5) plane (except when Θ_σ is very large, as in the bottom right panel). This implies that the secular dynamics of the 3/1 resonance on the plane $i = 0$ is dominated by the ν_5 resonance.

In panels like those of Fig. 11.7 the upper limit value of N is given by the condition $e \sim 1$. The curves that seem to exit from the top border of the panels therefore correspond to trajectories that reach $e \sim 1$ during their secular evolution and consequently hit the surface of the Sun. Conversely, the lower limit on the N axis is the value that, together with the considered value of $\Theta_\sigma \equiv J_\sigma$, identifies the separatrix of (11.5). As one sees in Fig. 11.4, the curves $J_\sigma =$ constant inside the resonance are defined only for sufficiently large eccentricity, namely for N larger than some lower threshold. Consequently, all the curves that seem to exit from the bottom border of the panels of Fig. 11.7 correspond to trajectories that hit the separatrix of the mean motion resonance during their secular evolution, and are therefore expected to be chaotic. Note however that in the present treatment of the problem, we have lost the notion of *modulation of the resonant amplitude* with the evolution of the secular angles, discussed in Section 9.4; in fact the action J_σ has been introduced for (11.5), where the secular angles do not appear, and has been afterwards considered as a constant.

The separatrices that surround the main libration island (see Fig. 11.7) can be represented in (a, e, i) space with a procedure equivalent to that discussed above for the separatrices of the Kozai resonance. First we define by σ_j^0 the value of σ_j for the stable equilibrium point at the center of the island; then we compute the values $\Theta_\nu^{(1)}$ and $\Theta_\nu^{(2)}$ of the intersections of the separatrices with the axis $\sigma_j = \sigma_j^0$. These values depend of course on Θ_σ and Θ_z. We obtain in this way a representation of the separatrices in $(\Theta_\sigma, \Theta_z, \Theta_\nu)$ space, which is then transformed to a representation in (a, e, i) space, as explained in Section 11.2.2.

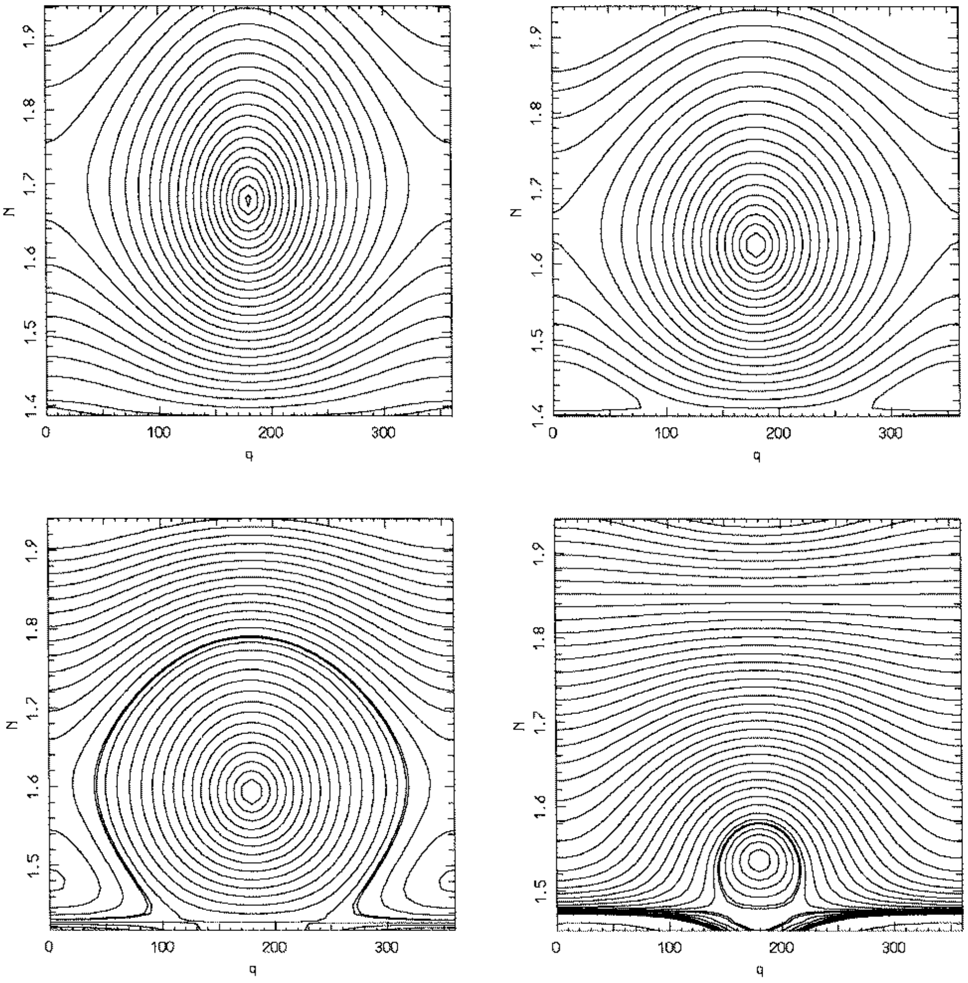


Figure 11.7: Level curves of (11.21) for the ν_5 resonance in the 3/1 commensurability. The panels are computed, from top left to bottom right, for increasing values of Θ_σ that are represented as pairs of dotted lines in Fig. 11.4. The label q stands for $\varpi - \varpi_5^*$, while $N = \Theta_\nu = \sqrt{a}(3 - \sqrt{1-e^2})$. At the center of the 3/1 resonance, the values $N = 1.4, 1.5, 1.6, 1.7, 1.8, 1.9$ correspond to eccentricity $e = 0.20, 0.55, 0.72, 0.84, 0.92, 0.97$, respectively. Reprinted from Fig. 7 of Moons and Morbidelli (1995), with permission from Academic Press.

In each region of circulation/libration of the secular resonance, one can introduce new Arnold action–angle variables $\Gamma^j_\sigma, \Gamma^j_z, \Gamma^j_\nu, \gamma^j_\sigma, \gamma^j_z, \gamma^j_\nu$ in order to eliminate the harmonic related to the $\dot{\vartheta}_\nu + g_5 = 0$ resonance.

This study can be done for all the proper frequencies g_j of planetary perihelia. For the mean motion resonances in the asteroid belt, one should at least study the resonances with g_5 and g_6, while in the Kuiper belt the most important resonance is that with g_8. Figure 11.4 shows the separatrices of the ν_5 and ν_6 resonances in the 3/1 mean motion resonance with Jupiter (see also Fig. 11.16 for the 3/2 mean motion resonance). As one sees, the two resonances are very large and completely overlap. As explained in Chapter 6, when a large overlap occurs, the dynamics is expected to become mostly chaotic. To check the extent of the chaotic region in a mean motion resonance due to the overlapping of the ν_5 and ν_6 secular resonances, one can construct a two-resonance model by retaining from (11.19) the harmonic of $\mathcal{K}_1$ with arguments $\sigma_5 = \vartheta_\nu + \varpi_5^*$ and $\sigma_6 = \vartheta_\nu + \varpi_6^*$. In practice, one considers the Hamiltonian

$$\mathcal{H}_{5,6} = g_5\Lambda_{g_5} + g_6\Lambda_{g_6} + \mathcal{K}_0(\Theta_\sigma, \Theta_z, \Theta_\nu) + \mathcal{K}_1^{5,6}(\Theta_\sigma, \Theta_z, \Theta_\nu, \sigma_5, \sigma_6) \ , \quad (11.22)$$

in the notation used for (11.21). This nonintegrable Hamiltonian must be studied numerically and the dynamics can be represented on the (σ_5, Θ_ν) plane, for instance through a section at $\sigma_6 = 0$. This is not a real Poincaré section, as the angle σ_6 can librate or circulate in both directions, but it is still suitable to distinguish regular from chaotic dynamics. To have a real Poincaré section, one should write $\sigma_6 = \sigma_5 + \varpi_6^* - \varpi_5^*$ and section on the value of $\varpi_6^* - \varpi_5^*$. Figure 11.8 shows the section on σ_6 of the dynamics of (11.22) for the 3/1 resonance, for the same values of Θ_σ used in Fig. 11.7, and still for $\Theta_z = 0$. As one sees, the difference with respect to Fig. 11.7 is striking, most of the phase space being covered by a chaotic region. This is a typical situation for all the main mean motion resonances with Jupiter in the asteroid belt, as first shown by Morbidelli and Moons (1993) and Moons and Morbidelli (1995). The extent of the chaotic region is the only difference among the resonances.

In the regular region that is not attained by the chaotic layer generated by the overlapping of the ν_5 and ν_6 resonances, one can introduce action–angle variables that remove at low order the dependence of the Hamiltonian on the angles ϖ_5^* and ϖ_6^*. This is done in two steps, first introducing the action–angle variables $\Gamma^5_\sigma, \Gamma^5_z, \Gamma^5_\nu, \gamma^5_\sigma, \gamma^5_z, \gamma^5_\nu$ that eliminate the harmonic in $\vartheta_\nu + \varpi_5^*$ and then introducing new action–angle variables $\Gamma^6_\sigma, \Gamma^6_z, \Gamma^6_\nu, \gamma^6_\sigma, \gamma^6_z, \gamma^6_\nu$ to eliminate the harmonic in $\gamma^5_\nu + \varpi_6^*$.

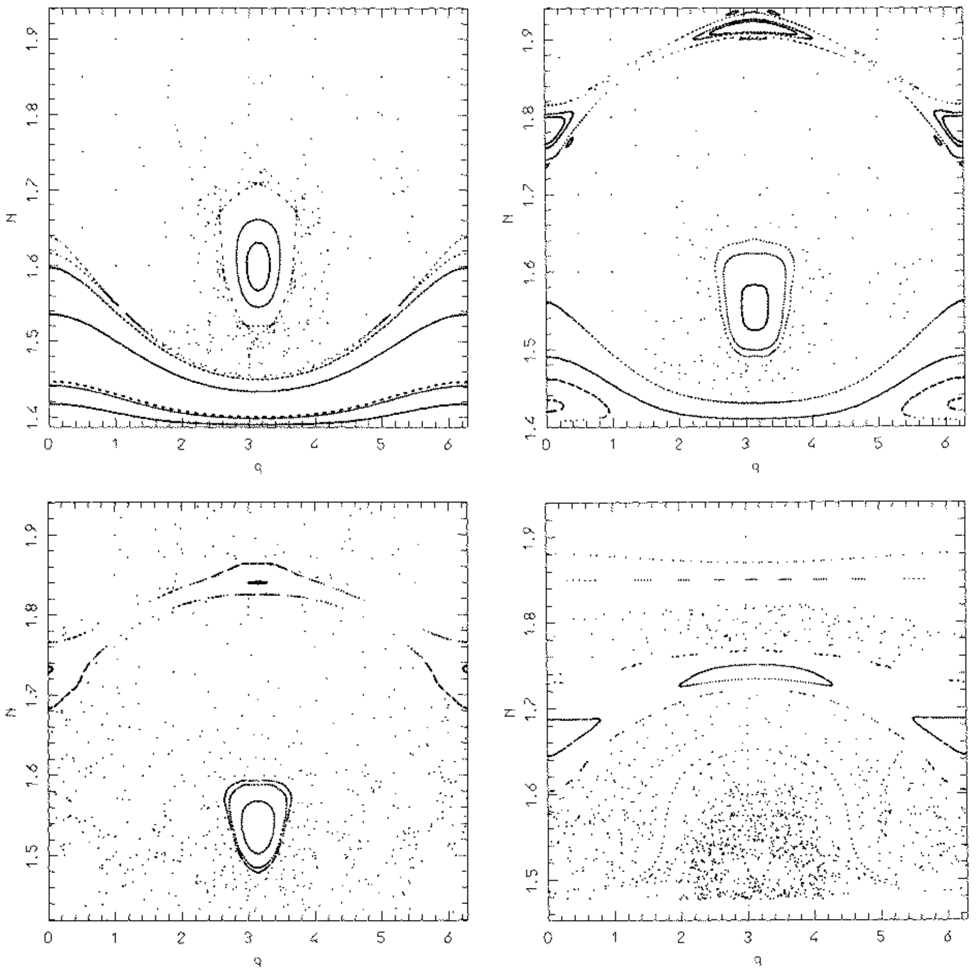


Figure 11.8: Sections of the dynamics of (11.22) computed at $\sigma_6 = 0$. The four panels correspond to the same values as Θ_σ of those of Fig. 11.7. The coordinates q and N are like in Fig. 11.7. See text for comments. Reprinted from Fig. 7 of Moons and Morbidelli (1995), with permission from Academic Press.

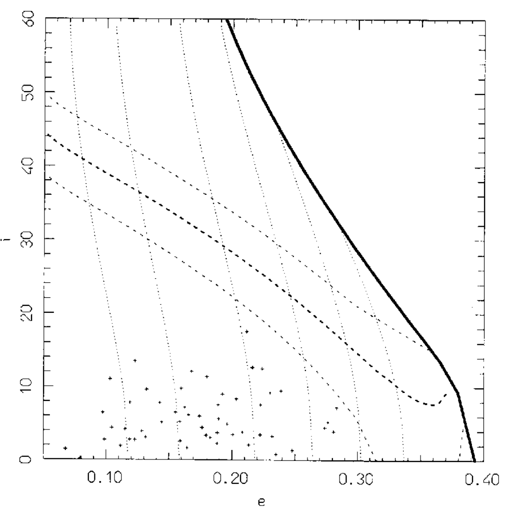


Figure 11.9: Secular dynamics at the center of the 3/2 resonance. The dotted curves denote the sets of (e, i) values that correspond to $\Theta' =$constant. The bold solid curve marks the leftmost separatrix of the Kozai resonance. The bold dashed and the light dashed curves denote respectively the center and the separatrices of the ν_{16} secular resonance. The dots represent the osculating eccentricity and inclination of the Hildas. Reprinted from Fig. 14 of Morbidelli and Moons (1993), with permission from Academic Press.

11.2.4 Nodal secular resonances

Figures 11.10, 11.2 and 11.17 show the intersection with the plane $i = 0$ of the location of the resonance $\dot{\vartheta}_z + \dot{\vartheta}_\nu + s_6 = 0$ (also called ν_{16}) in the 2/1, 3/1 and 3/2 mean motion resonances in the asteroid belt. Analogously, Fig. 11.18 shows the location of the resonance $\dot{\vartheta}_z + \dot{\vartheta}_\nu + s_8 = 0$ (also called ν_{18}) in the 2/3 resonance with Neptune in the Kuiper belt.

To study a resonance of type $\dot{\vartheta}_z + \dot{\vartheta}_\nu + s_6 \sim 0$ we isolate in (11.19) the harmonic of $\mathcal{K}_1$ with argument $q_j = \vartheta_z + \vartheta_\nu + \Omega_j^*$, and consider the integrable model

$$\mathcal{H}_j = s_j \Lambda_{s_j} + \mathcal{K}_0(\Theta_\sigma, \Theta_z, \Theta_\nu) + \mathcal{K}_1^j(\Theta_\sigma, \Theta_z, \Theta_\nu, q_j) \ . \tag{11.23}$$

The function $\mathcal{K}_1^j$ is the average of $\mathcal{K}_1$ over all the angles except $\vartheta_z + \vartheta_\nu + \Omega_j^*$. The dynamics of (11.23) preserves the value of Θ_σ and $\Theta' = \Theta_z - \Theta_\nu$, while Θ_z (and consequently Θ_ν) evolves with the circulation/libration of q_j.

The condition $\Theta' =$constant can be transformed into a relationship between i and e on the surface defined by the value of Θ_σ and by $\sigma = \sigma_0$, with the following procedure: for each value of Θ_z and $\Theta_\nu = \Theta_z - \Theta'$, one first computes the value of J_z that corresponds to Θ_z when $\psi_z = \psi_z^0$, and then transforms the resulting values of $(J_z, J_\nu \equiv \Theta_\nu)$ into (e, i) as explained in Section 11.2.2 for the separatrices of the Kozai resonance. Figure 11.9 shows some curves $\Theta' =$constant on the (e, i) plane defined by $\Theta_\sigma = 0$, in the case of the 3/2 mean motion resonance with Jupiter. On these curves the eccentricity is almost constant, showing that the nodal secular resonance acts mainly on the asteroid's inclination, as expected.

The secular evolution of Θ_z, q_j can be represented with the level curves of (11.23), Θ_σ and Θ' playing the role of parameters, analogously to what has been explained in Section 11.2.3 for the perihelion resonances. Also, one can compute the extreme values $\Theta_z^{(1)}$ and $\Theta_z^{(2)}$ assumed by Θ_z on the separatrices of the considered secular resonance. The set of values $(\Theta_z^{(1)}, \Theta_\nu = \Theta_z^{(1)} - \Theta')$ and $(\Theta_z^{(2)}, \Theta_\nu = \Theta_z^{(2)} - \Theta')$ can be transformed into values of e and i on the surface defined by the value of Θ_σ and by $\sigma = \sigma_0$, thus obtaining a representation of the separatrices of the secular resonance like that of Fig. 11.9.

As usual, new action–angle variables can be introduced, in each libration/circulation region of the secular resonance, in order to remove the considered resonant harmonic.

11.2.5 Three-body resonances

The existence and importance of three-body resonances inside two-body mean motion resonances has been pointed out only recently by Ferraz-Mello and collaborators in a series of works (Ferraz-Mello, 1996; Ferraz-Mello *et al.*, 1997; Michtchenko and Ferraz-Mello, 1997; Nesvorný and Ferraz-Mello, 1997b; Ferraz-Mello *et al.*, 1998a). These are commensurabilities occurring between the frequency $\dot{\psi}_\sigma$ and the frequency of a quasi-resonant combination of the mean longitudes of two planets, mainly $2\dot{\lambda}_{\rm J} - 5\dot{\lambda}_{\rm S}$ for the asteroid belt and $\dot{\lambda}_{\rm U} - 2\dot{\lambda}_{\rm N}$ for the Kuiper belt, where $\dot{\lambda}_{\rm J}, \dot{\lambda}_{\rm S}, \dot{\lambda}_{\rm U}, \dot{\lambda}_{\rm N}$ denote the *mean* mean motions of Jupiter, Saturn, Uranus and Neptune, respectively. These resonances can be seen as a consequence of the overlap between mean motion resonances with two different planets, in the case where one resonance is much stronger than the other. In this sense, the use of the terminology "three-body resonance" is quite an abuse, and in fact we will show below that the coefficients of the related harmonics are linear in the planetary mass, while in the case of the "real" three-body resonances the coefficients are quadratic (see Chapter 10).

To clarify, consider the specific case of the 2/1 mean motion resonance

with Jupiter; because $2\dot{\lambda}_{\rm J} - 5\dot{\lambda}_{\rm S}$ is small (the corresponding period is about 880 years) the asteroid is also close to the 5/1 mean motion resonance with Saturn. Therefore, in the construction of the normal form for the 2/1 resonance with Jupiter, one cannot average out the harmonics of (2.38) with arguments $m(5\lambda_{\rm S} - \lambda + \cdots)$, where m is an integer number and the dots stand for the correct combinations of the secular angles that satisfy the D'Alembert rules. These harmonics have a coefficient that is proportional to $\varepsilon_{\rm S} e^{\alpha} e_{\rm S}^{\alpha_{\rm S}} i^{\beta} i_{\rm S}^{\beta_{\rm S}}$ where $\varepsilon_{\rm S}$ is the mass of Saturn relative to the Sun, $e, i, e_{\rm S}, i_{\rm S}$ are the eccentricity, the inclination of the asteroid and of Saturn and $\alpha + \alpha_{\rm S} + \beta + \beta_{\rm S} \geq 4m$. Because the eccentricity and the inclination of Saturn are small, the harmonics with the largest coefficients are those with $\alpha_{\rm S} = \beta_{\rm S} = 0$. If the asteroid's inclination is moderate, then, for each m, the dominating harmonic is that with argument $m(5\lambda_{\rm S} - \lambda - 4\varpi)$. Its coefficient is proportional to $\varepsilon_{\rm S} e^{4m}$. Of course, one can rewrite:

$$m(5\lambda_{\rm S}-\lambda-4\varpi) = m(2\lambda_{\rm J}-\lambda-\varpi)+m(5\lambda_{\rm S}-2\lambda_{\rm J})-3m\varpi = m\sigma+m\varphi_{\rm JS}-3m\varpi \tag{11.24}$$

where $\varphi_{\rm JS}$ denotes the so-called *great inequality angle* $5\lambda_{\rm S} - 2\lambda_{\rm J}$.

Because these dominating harmonics have coefficients that are much smaller than that of the main harmonic associated with the 2/1 resonance with Jupiter, their effect can be studied with the hierarchical perturbation approach that characterizes this chapter (the opposite case – where the resonances with the two planets have comparable strength – has been treated in Section 9.2.2). One first introduces the variables (11.6) for the 2/1 resonance with Jupiter. Then, each harmonic with argument (11.24) gives rise to a series of harmonics with arguments

$$n\psi_\sigma + m\varphi_{\rm JS} + 3m\psi_\nu \tag{11.25}$$

with $n \in \mathbf{Z}$ and coefficients exponentially decaying with $|n|$. Therefore, for every resonance of type $n\dot{\psi}_\sigma \sim m\dot{\varphi}_{\rm JS}$ there are corresponding harmonics with nonnegligible coefficients in the perturbation function.

Because in general these terms are nevertheless quite small, in the process of successive elimination of harmonics, the study of three-body resonances logically comes after the consideration of the main secular terms. With the introduction of action–angle variables $\Gamma_\sigma^j, \Gamma_z^j, \Gamma_\nu^j, \gamma_\sigma^j, \gamma_z^j, \gamma_\nu^j$ for the regular trajectories outside a perihelion secular resonance with a planetary frequency g_j, each harmonic (11.25) generates a multiplet of harmonics with arguments

$$n\gamma_\sigma^j + m\varphi_{\rm JS} + (3m+k)\gamma_\nu^j + k\varpi_j^* \ , \quad k \in \mathbf{Z} \tag{11.26}$$

where γ_ψ^j and γ_ν^j are the new angles, respectively "close" to ψ_σ and $\psi_\nu \sim -\varpi$. This can be more intuitively understood as follows: the use of the new variables

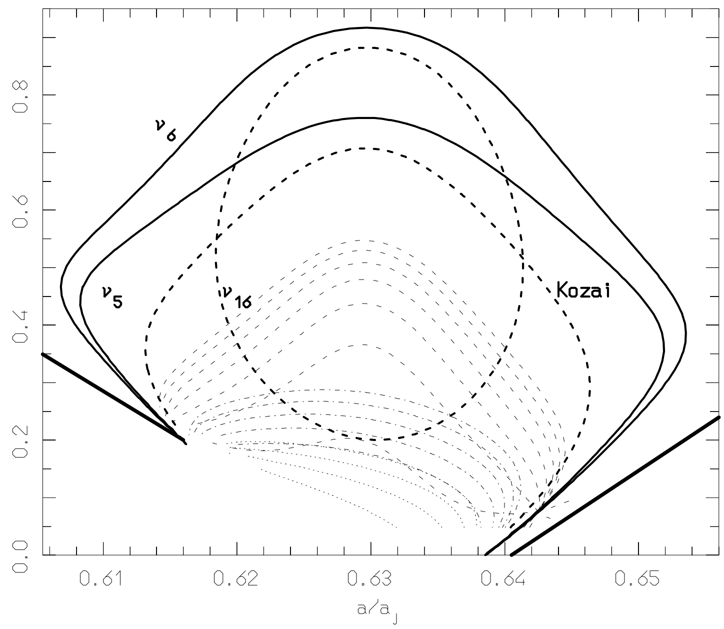


Figure 11.10: The location of the main resonances inside the 2/1 commensurability. The two thick lines denote the separatrices of the 2/1 resonance, plotted for $\sigma = 0$ degrees. The bold solid and dashed curves are like in Fig. 11.2; the light dotted curves refer to the secondary resonances $\dot{\psi}_\sigma + n(\dot{\psi}_\nu + g_5) = 0$ ($n = 2, \ldots, 5$); the light dashed curves denote the location of the three-body resonances $\dot{\psi}_\sigma + 2\dot{\varphi}_{\rm JS} - (6-k)\dot{\varpi} - kg_5$ with k ranging from 0 to 6 with decreasing eccentricity; the light dashed dotted curves mark the resonances $\dot{\psi}_\sigma + \dot{\varphi}_{\rm JS} - (3-k)\dot{\varpi} - kg_5$ with k ranging from 0 to -4 with increasing eccentricity. Remember that $\dot{\psi}_\sigma < 0$ for interior resonances.

is essentially equivalent to considering that the asteroid's eccentricity has a secular oscillation of type $e = e_0 + \delta\cos(\varpi - \varpi_j^*)$; by substituting this expression into the term $e^{4m}\cos(n\psi_\sigma + m\varphi_{\rm JS} - 3m\varpi)$ one obtains the harmonics with arguments $n\gamma_\sigma + m\varphi_{\rm JS} - (3m+k)\varpi + k\varpi_j^*$ with $-4m \le k \le 4m$, which correspond to those in (11.26). As a matter of fact, these harmonics are already present in the direct perturbation given by Saturn, with coefficients of order $\varepsilon_{\rm S} e_0^{|4m+k|} e_{\rm S}^{|k|}$; the secular oscillation of the asteroid's eccentricity generates them again with coefficients not smaller than quantities of order $\varepsilon_{\rm S} e_0^{4m-|k|}\delta^{|k|}$. As the amplitude δ is in general much larger than Saturn's eccentricity $e_{\rm S}$, the indirect effect produced by the asteroid's secular eccentricity oscillation largely dominates over the direct effect of Saturn's eccentricity.

Figure 11.10 shows the location of the resonances with critical angles $\psi_\sigma + 2\varphi_{\rm JS} - (6-k)\varpi - k\varpi_5^*$ and $\psi_\sigma + \varphi_{\rm JS} - (3-k)\varpi - k\varpi_5^*$ on the plane $i = 0$ for the 2/1 resonance with Jupiter, computed using the frequencies given by (11.9). As one sees, these resonances form a very dense web over almost the entire 2/1 resonant region. Although individually weak, the cumulative effect of these resonances may be important to make chaotic those regions which are not affected by notable secondary and secular resonances.

Up to now, the effect of three-body resonances inside two-body mean motion resonances has never been studied in detail with analytic or semianalytic models. Only through numerical integrations has it been shown (see for instance Ferraz-Mello *et al.*, 1998a) that the dynamics may turn from regular to chaotic when the effect of Saturn is introduced, and that the strength of chaos depends on the frequency of the great inequality angle (which can be varied by slightly changing Saturn's orbit). On the other hand, we know from Chapter 6 that it is very hard to analytically conclude whether a given system is quasi-regular or completely chaotic, when a large number of thin high-order resonances are involved. In these cases, for a clear understanding of the real dynamical structure of the system it is necessary to resort to numerical explorations, for instance using the tools described in Chapter 5.

11.3 The major resonances in the asteroid belt

In this section we discuss our present understanding of the dynamics inside the most important mean motion resonances with Jupiter in the asteroid belt. For a historical review of the progress of this understanding, we recommend the paper by Moons (1997).

11.3.1 The 3/1 resonance

The 3/1 mean motion resonance with Jupiter at $a \sim 2.5$ AU corresponds to one of the most evident Kirkwood gaps in the asteroid belt.

The secondary resonances are not important for the dynamics inside the 3/1 resonance, as they are present only at small eccentricity ($e < 0.1$, see Fig. 11.2). Therefore, the origin of the gap must be searched for in the secular dynamics. Despite the fact that the exact location of the ν_5 resonance is at large eccentricity ($e \sim 0.8$, see Fig. 11.4), this resonance dominates the secular dynamics for all values of e (see Fig. 11.7). As a consequence, the 3/1 resonant asteroids have large secular oscillations of the eccentricity, which force them to temporarily cross the orbits of the inner planets.

Wisdom (1982) was the first to put this phenomenon in evidence, (see

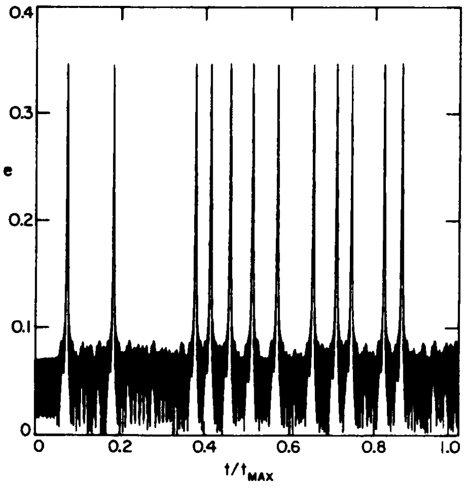


Figure 11.11: The time evolution of the eccentricity of an asteroid in the 3/1 resonance, in the framework of the planar elliptic restricted problem, according to Wisdom's symplectic mapping. Reprinted from Fig. 13 of Wisdom (1982), with permission from the American Astronomical Society.

Fig. 11.11) with numerical simulations performed using a mapping which can be thought of as the ancestor of the *mixed variable symplectic integrator* (Wisdom and Holman, 1991). Later, Wisdom (1985) explained the observed phenomenon with a semianalytic theory. His approach was similar to that discussed in Section 11.2.3, but took into account also the secular modulation of the separatrices of the 3/1 resonance and was limited, being based on series expansions, to low–moderate values of the eccentricity (see Henrard and Caranicolas, 1990, for a commented review of Wisdom's theory). Figure 11.12 (from Wisdom, 1985) shows the secular motion of $\bar{x} = e\cos(\varpi - \varpi_5^*), \bar{y} = e\sin(\varpi - \varpi_5^*)$, obtained by averaging over the motion of σ. The shaded area denotes the separatrix of the 3/1 resonance; σ circulates in the region enclosed by the shaded area, and librates otherwise. As one sees, almost all the orbits that are in the 3/1 resonance (i.e. outside of the shaded area) are forced by the secular dynamics to exceed $e = 0.3$, a threshold value for which close encounters with Mars may occur. The trajectories of Fig. 11.12 that hit the shaded area are all expected to be chaotic, so that they can start their excursion to large eccentricity at random time, as shown in the numerical integration of Fig. 11.11. This kind of behavior is also visible in the top right panel of Fig. 11.7, computed following the recipe of

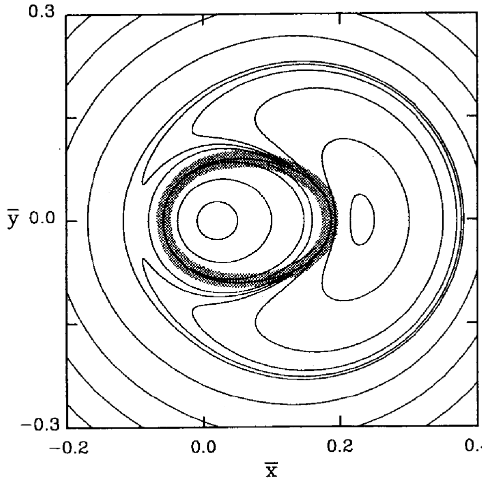


Figure 11.12: The secular motion in the 3/1 resonance, for a specific value of the averaged Hamiltonian of the planar elliptic three-body problem. The coordinates are $\bar{x} = e\cos(\varpi - \varpi_5^*), \bar{y} = e\sin(\varpi - \varpi_5^*)$, where e and ϖ are the values of the eccentricity and perihelion longitude, averaged over the motion of σ. The shaded area represents the separatrix of the 3/1 resonance. The angle σ circulates for the trajectories enclosed by the shaded area, and librates for all other trajectories. Reprinted from Fig. 5 of Wisdom (1985), with permission of Academic Press.

Section 11.2.3; in that panel, the bottom border represents the separatrix of the 3/1 resonance, and therefore corresponds to the shaded area of Fig. 11.12, while the island of libration around $q = \varpi - \varpi_5^* = 0$ at small N corresponds to the banana-shaped curves on the right-hand side of Fig. 11.12.

For orbits with smaller amplitude of libration inside the 3/1 resonance (top left panel in Fig. 11.7), the island of secular libration around $\varpi - \varpi_5^* = 0$ disappears, as first shown by Henrard and Caranicolas (1990). However the orbits that have the lowest possible eccentricity when $\varpi - \varpi_5^* = 180°$ still reach eccentricity larger than ~ 0.3 (i.e. $N \geq 1.42$) when $\varpi - \varpi_5^*$ precesses down to $0°$. Despite the fact that this evolution in itself is not chaotic, the acquired large eccentricity is sufficient to allow encounters with Mars. Planetary encounters give random kicks to the asteroid's semimajor axis, and can eventually remove it from the 3/1 resonance. As a consequence of this interplay between secular dynamics and Martian encounters, starting with a uniform distribution of particles a gap could open corresponding to the 3/1 resonance in an

estimated time of several 100 My (Wisdom, 1983; Wetherill, 1975). However, this is not the real way by which the resonance gets depleted of objects, as we will discuss below.

In Fig. 11.7, the big island of libration around $\varpi - \varpi_5^* = 180°$ is what properly corresponds to the ν_5 resonance. The trajectories close to the bottom of the two top panels are outside of this big island, and therefore reach only moderate (although Mars-crossing) values of the eccentricity during the circulation of $\varpi - \varpi_5^*$ or its libration around 0 degrees. On the contrary, the trajectories with initially larger eccentricity ($e \sim 0.45$ when $\varpi - \varpi_5^* = 180°$) follow a libration cycle around $\varpi - \varpi_5^* = 180°$ and reach an eccentricity that is close to unity. For larger amplitude of libration (bottom left panel of Fig. 11.7) the big island of libration touches the separatrix of the 3/1 resonance (the bottom level of the panel) and thus also the orbits with the lowest inital eccentricity are forced to acquire an eccentricity of order 0.8–0.9 when $\varpi - \varpi_5^* = 180°$. This property was first shown by Ferraz-Mello and Klafke (1991) by computing Poincaré sections of the averaged planar elliptic three-body problem, which has the form (11.13) with $S_z = 0$. The traces of the separatrices of the big island of libration on the (a, e) plane at $\sigma = 90°$ are shown in Fig. 11.4.

The same figure also shows the location and the separatrices of the ν_6 resonance which, by itself, gives a secular dynamics very similar to that shown in Fig. 11.7. When both resonances are taken simulatenously into account, as explained in Section 11.2.3, the secular dynamics changes completely (Fig. 11.8). A chaotic region dominates almost the entire phase space and extends to eccentricity equal to 1 (top borders of the panels). Apart from small islands embedded in the chaotic region, only the trajectories with low amplitude of libration in the 3/1 resonance and low eccentricity (the invariant tori at the bottom of the two top panels of Fig. 11.8) still have a regular dynamics. Remember however that these trajectories periodically reach a Mars-crossing eccentricity; the encounters with Mars give impulse changes to the semimajor axis and the eccentricity of the asteroid, which – although generally very small – effectively modify the asteroid's amplitude of libration in the 3/1 resonance. At large amplitude of libration, the chaotic region generated by the overlapping of the ν_5 and ν_6 resonances extends to all eccentricities (bottom panels of Fig. 11.8), and the asteroid can therefore rapidly and chaotically evolve to very large eccentricity.

This combination of large-scale chaos of the secular dynamics and weak Martian encounters well explains the behavior of 3/1 resonant asteroids observed in numerical integrations of the full equations of motion. As Fig. 11.13 shows, if the eccentricity and amplitude of libration are both initially small, the asteroid may spend ~ 1 My in a low-eccentricity mode, oscillating be-

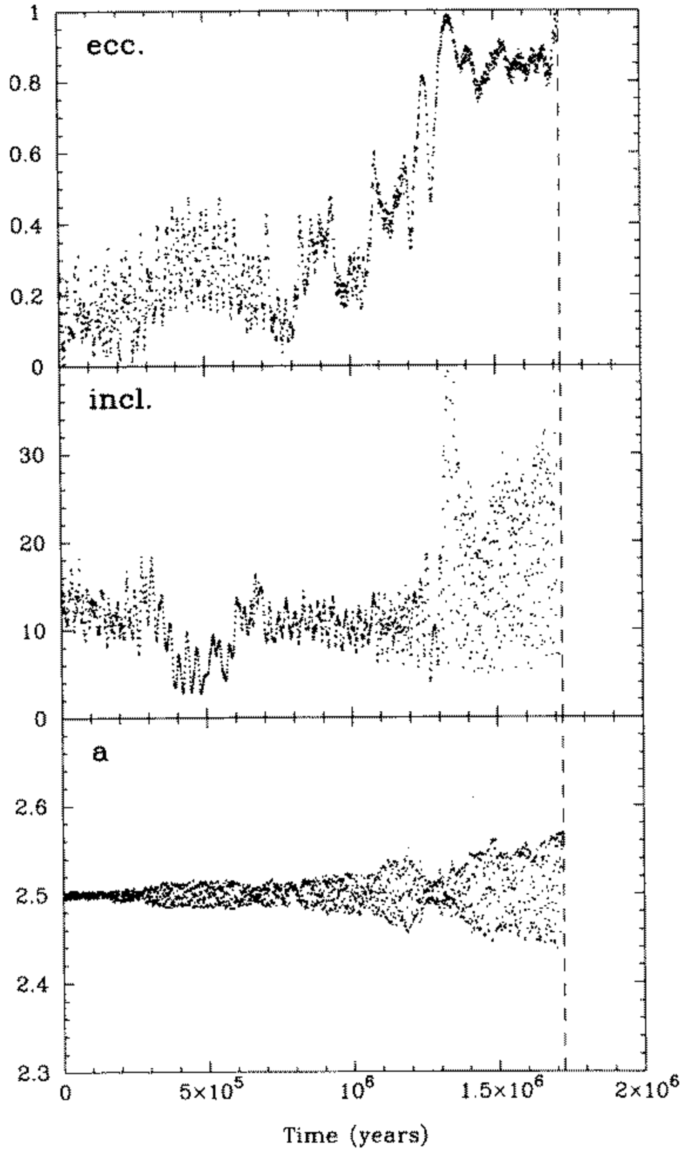


Figure 11.13: The typical evolution of eccentricity, inclination and semimajor axis of a body initially in the 3/1 resonance with Jupiter, as resulting from a numerical integration of the full equations of motion, in a model that includes all the planets of the Solar System. The body eventually hits the Sun. Reprinted from Fig. 9 of Farinella *et al.* (1993a), with permission from Kluwer Academic Publishers.

tween $e \sim 0$–0.1 and $e \sim 0.3$–0.4 on a timescale of order 10^4 y. Meanwhile the amplitude of libration of the semimajor axis is modified by Martian encounters. The asteroid eventually enters in the large chaotic region of the secular dynamics and its eccentricity increases very rapidly to values close to unity. Encounters with the Earth and Venus are now possible and frequent. For their larger masses, these two planets are much more effective than Mars in changing the asteroid's semimajor axis; if a "strong" encounter occurs, the asteroid is removed from the 3/1 resonance, otherwise – as in Fig. 11.13 – the effect of the planetary encounters results only in a continuous modification of the asteroid's amplitude of libration, while its eccentricity continues its random evolution due to the chaotic secular dynamics. In this case, the typical endstate is a collision with the Sun. Note that, during the evolution, the asteroid's inclination has also largely changed. This is the result of the existence, inside of the 3/1 resonance, of secular resonances that act on the evolution of the asteroid's inclination, as the Kozai resonance, the ν_{16} resonance and the $\dot{\vartheta}_\nu + g_6 + 2\dot{\vartheta}_z = 0$ resonance, whose locations are shown on the (a, e) plane in Fig. 11.2.

The dynamical evolution shown in Fig. 11.13 results in the rapid formation of a gap in the asteroid distribution corresponding to the 3/1 resonance. In fact, the numerical integration of a statistically significant number of particles (Gladman *et al.*, 1997) shows that the median lifetime of a population initially in the 3/1 resonance is 2 My; less than 10% of the population survives for more than 10 My, but these surviving bodies have all been transported to $a < 2$ AU by repeated strong encounters with the Earth or Venus, so that the resonance is completely emptied in a few million years. This timescale is much shorter than the ~ 100 My timescale predicted by Wisdom's scenario.

The same happens in many other major mean motion resonances with Jupiter that are associated with a Kirkwood gap. Moons and Morbidelli (1995) have shown that also the 4/1, 5/2 and 7/3 resonances are dominated by the chaotic region generated by the overlapping of the ν_5 and ν_6 resonances, so that also the asteroids in these mean motion resonances can reach very large eccentricities on a "short" timescale. Gladman *et al.* (1997) have computed that the median lifetime of the particles in the 5/2 and 7/3 resonances is respectively equal to 0.6 and 19 My.

For their efficiency in pumping the eccentricity up to Earth-crossing values, the major mean motion resonances with Jupiter – in particular the 3/1 resonance – are considered to be the most important transport routes for supplying new objects to the near-Earth asteroid population. In the generally accepted scenario, the resonances are continuously refilled with asteroids from the main belt population at low eccentricity, because of collisions (Farinella *et*

al., 1993b, Menichella *et al.*, 1996) and of the slow action of a weak nongravitational force (the so-called *Yarkovsky effect*[4]). Because the chaotic region in the 3/1 resonance allows the eccentricity to reach unity on a "short" timescale, most of the resonant asteroids are rapidly eliminated by a collision with the Sun (70% of the bodies in the Gladman *et al.* integrations), and therefore – luckily for us! – do not significantly contribute to the maintenance of the NEA population.

What would happen in the 3/1 resonance if Mars did not exist? The dynamics of the trajectories that appear regular in the top panels of Fig. 11.8 has never been studied in detail, taking into account additional harmonics and degrees of freedom of the secular dynamics. As Fig. 11.2 shows, the resonance $\dot{\vartheta}_\nu + g_6 + 2\dot{\vartheta}_z = 0$ crosses this region, and may potentially destabilize the motion. Also two three-body resonances cross this region, but their harmonics should be of very large eccentricity order, because the lowest order resonance with Saturn that may interfere with the 3/1 resonance with Jupiter is the 15/2 (order 13 in e). A numerical integration of the same particles as Gladman *et al.* (1997), but without the presence of Mars, shows that the population in the 3/1 resonance decays much more slowly. After 10 My, 15% of the particles still survive in the resonance, while in Gladman *et al.* integrations the 3/1 resonant region is completely cleared on this timescale.

11.3.2 The 2/1 resonance

The 2/1 mean motion resonance with Jupiter is associated with the Kirkwood gap in the asteroid belt of largest width in the semimajor axis. As a consequence, it was the first resonance to be studied in detail, since the work of Poincaré. Curiously, it is also the resonance for which it is most difficult to explain the absence of asteroids, and a quite satisfactory understanding has been achieved only very recently.

The dynamical structure of the 2/1 resonance has been revisited by Moons *et al.* (1998) by applying, for the first time without simplifications, the approach explained throughout Section 11.2. The computation of the location and the separatrices of the various resonances has been done for several values of J_σ, and the results have then been collectively represented on the (a, e) plane, for various values of the inclination (Fig. 11.14). In each panel, the

[4]This effect is produced by the thermal re-emission of the asteroid. Because of the diurnal rotation, or of the obliquity of the spin axis, the hottest part of the body's surface emits infrared photons in a direction that is offset with respect to the Sun–body direction. This results in a weak acceleration (positive or negative) of the body along its orbit, namely in a slow drift of its semimajor axis. This phenomenon is directly observed for the Earth's artificial satellites. See Farinella and Vokrouhlický (1998); Bottke *et al.* (2000a)

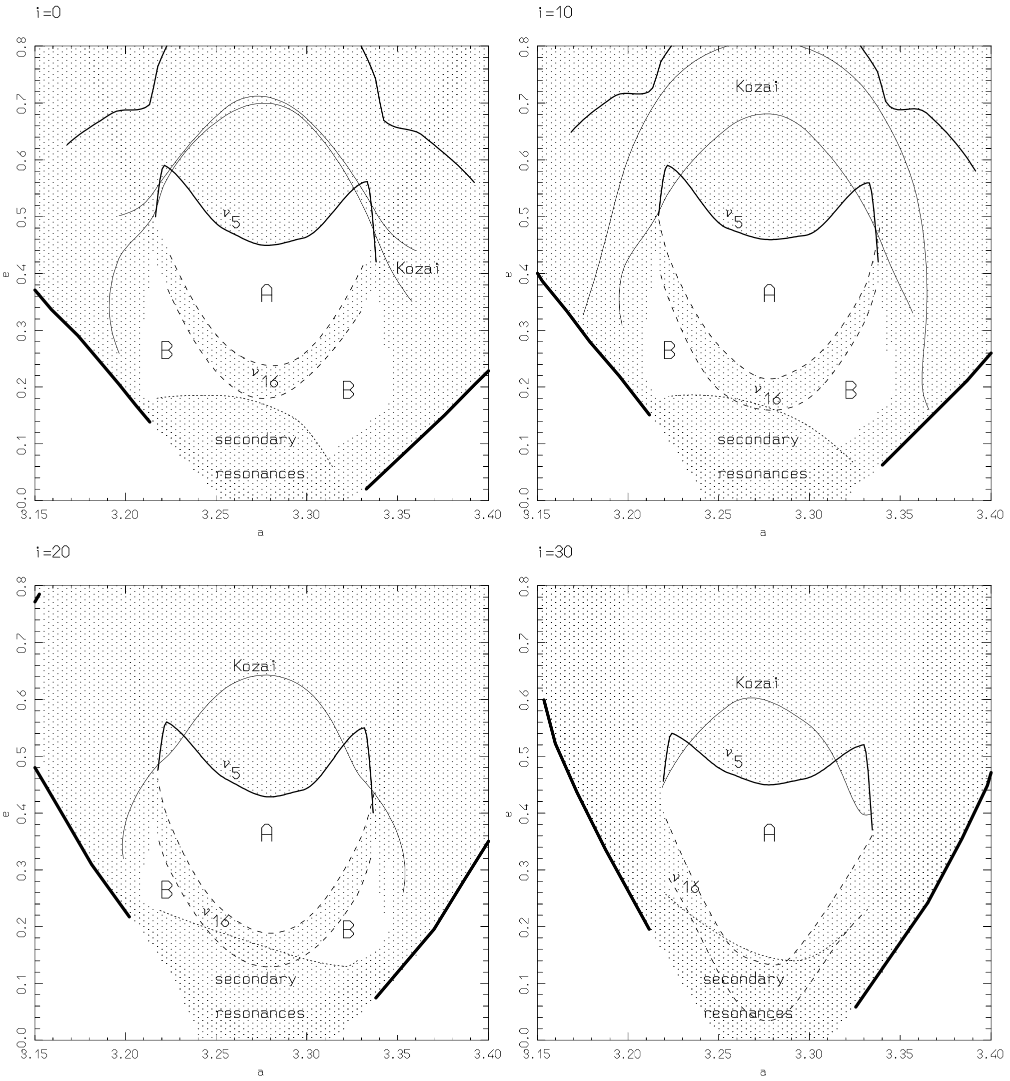


Figure 11.14: Strongly chaotic regions (dotted) and quasi-regular islands (white areas labelled A or B) in the 2/1 resonance with Jupiter at different inclinations. The two thick lines on the sides denote the separatrices of the 2/1 mean motion resonance; the dotted curve marks the upper limit of the secondary resonances; the dashed, light solid and bold solid curves denote the separatrices of the ν_{16}, Kozai and ν_5 resonances, respectively. See text for comments. Reprinted from Fig. 1 of Moons *et al.* (1998), with permission from Academic Press.

two thick lines on the sides denote the separatrices of the 2/1 mean motion resonance, represented for $\sigma = 0$. Note that, with increasing inclination, the region bounded by the separatrices of the 2/1 commensurability moderately shrinks. The separatrices are in fact computed from the Hamiltonian (11.5), which depends parametrically on the inclination through S_z. In between the separatrices, the points color the regions concerned by one or more secular or secondary resonances discussed below. This is aimed at highlighting the regions where strong (chaotic) variations of the orbital elements are expected.

The light solid curves in Fig. 11.14 refer to the separatrices of the Kozai resonance. The width of the Kozai resonance increases with the inclination, because the coefficient of the Kozai harmonic is proportional to $e^2 i^2$. A view of the resonance in the (e, i) plane at the center of the 2/1 commensurability is given in Fig. 11.6. At $i \geq 20°$ the upper Kozai separatrix is outside of the eccentricity range covered by Fig. 11.14 for most values of the semimajor axis, so that only the lower separatrix is plotted. The bold solid curves denote the separatrices of the ν_5 resonance. The angle $\varpi - \varpi_5^*$ librates around 0 if the semimajor axis oscillates in the range 3.22–3.33 AU and it librates around 180 degrees for orbits with larger amplitude of libration in the 2/1 resonance. The ν_6 secular resonance is also present (Morbidelli and Moons, 1993): it is located above the ν_5 and partially overlaps with the former, but its separatrices are not represented in Fig. 11.14, for simplicity. The Kozai resonance stongly interacts with the ν_5 and ν_6 resonances; partial overlapping occurs at all inclinations and this is expected to give rise – as for the 3/1 resonance – to a large chaotic region, where the eccentricity of the asteroid can grow to unity. Such chaos covers the region at large eccentricity at the center of the 2/1 resonance, while – at large libration amplitudes – it extends to all eccentricities, following the location of the above mentioned secular resonances (see Fig. 11.10). Therefore, considering only the Kozai, ν_5 and ν_6 resonances, the region with small to moderate eccentricity at the center of the 2/1 resonance would be regular. In contrast to the 3/1 resonance case, here one cannot invoke Martian encounters to increase the amplitude of libration of the asteroids initially in this region, because the eccentricity does not reach large enough values to cross the orbit of Mars (remember that the 2/1 resonance is much further from Mars than the 3/1 resonance).

Explaining the absence of asteroids in the small to moderate eccentricity region requires further investigation of the dynamics. This region is crossed by the ν_{16} resonance, whose separatrices are represented in Fig. 11.14 by dashed curves. The ν_{16} resonance forces oscillations of the inclination of about 15–20 degrees in amplitude. Moreover, at lower eccentricity, there is the complex of secondary resonances: the dotted curve marks the upper limit in eccentricity of

the region concerned by the presence of secondary resonances whose orders in eccentricity or inclination are no larger than 5. Therefore, accounting also the ν_{16} resonance and for the secondary resonances, only the white regions labelled by A and B in Fig 11.14 are expected to host regular dynamics. In a planar problem, the bodies in the region of secondary resonances, although chaotic, could not escape because they should cross the stable A and B regions (see Fig. 11.3). In a complete problem, however, the secondary resonances of type (11.10) and (11.12) pump the inclination of the asteroid. At larger inclination the ν_{16} resonance overlaps with the secondary resonances (bottom panels of Fig. 11.14), crunching the B region: the way is therefore open to escape to large eccentricity, through the chaotic region generated by the Kozai, ν_5 and ν_6 resonances. This tortuous path was first shown numerically by Wisdom (1987) and later described in detail by Henrard *et al.* (1995), through a series of numerical experiments.

Nevertheless, neither the secular nor the secondary resonances fully explain why the A and the B regions are not densely inhabited by asteroids. These regions, however, are not really regular, as shown by Franklin (1994) and Ferraz-Mello (1994) by computing Lyapunov exponents for a limited set of orbits. A few years later, Nesvorný and Ferraz-Mello (1997b) carried the first detailed numerical survey of the dynamical structure of the 2/1 resonance. They have integrated for several 10^5 y a large number of test particles initially placed on a regular grid of the (a, e) plane at $i = \sigma = \varpi - \varpi_J = \Omega - \Omega_J = 0$, the subscript J denoting Jupiter's angles. For each test particle, they measured the change over time of the frequency of ϖ. As discussed in Section 5.3, the change of the frequency is a measure of the degree of chaos of the orbit. The results have been color coded in Fig. 11.15. Red denotes the strongest chaotic regions, while dark blue indicates the initial conditions for which the change of the frequency is below the dectection limit. The correspondence between the left panel of Fig. 11.15 and the top left panel of Fig. 11.14, for what concerns the location of the most chaotic regions, is remarkable. In addition, the power of the numerical technique reveals that the A and B regions are characterized by an extended weak chaos – as indicated by a almost uniform light blue color – and not by the Nekhoroshev structure discussed in Chapter 6. For this reason, one should expect that the asteroids initially in the A or B region exhibit slow chaotic diffusion on the (a, e) plane, and may eventually join – after a long timescale – the strong chaotic regions. In fact, the long-term numerical integrations done by Nesvorný (private communication) show that the bodies initially in the A and B region have a median lifetime exceeding 1 Gy. Mitchenko and Ferraz-Mello (1997) and Nesvorný and Ferraz-Mello (1997b) suggested that the weak extended chaos of the A and B regions is

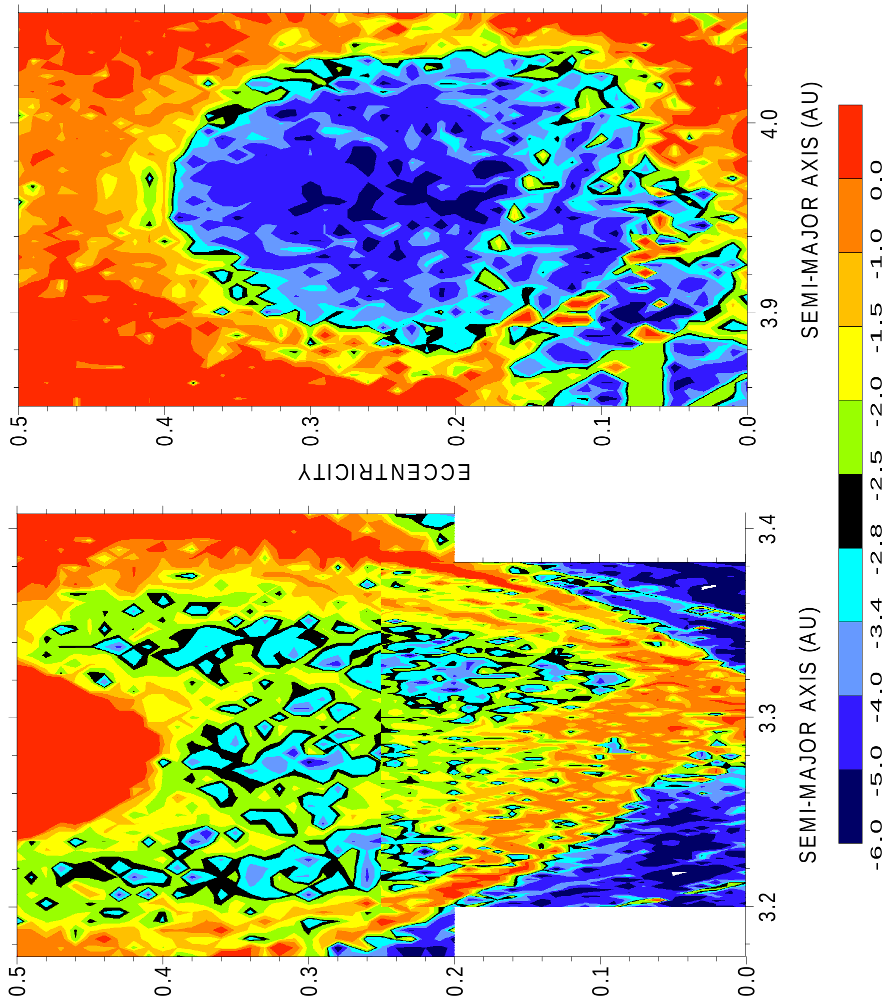


Figure 11.15: The dynamical structure of the 2/1 and 3/2 resonances (left and right panels respectively). The color scale codes the logarithm of the relative change of the frequency of ϖ over 200,000 y, and should be considered as an indicator of chaos. Reprinted from Fig. 2 of Ferraz-Mello *et al.* (1998b), with permission from the Astronomical Society of the Pacific.

caused by resonances between the libration period of σ and the period of the great inequality angle $\varphi_{\rm JS}$ (see Section 11.2.5). For an indirect proof, they slightly changed the orbit of Saturn, in order to reduce the period of the great inequality angle by a factor of 2, showing that the degree of chaos of these regions is strongly enhanced. As Fig. 11.10 shows, the A and B regions are in fact crossed by the complex of "three body resonances" with critical angles (in the notation of Section 11.2.5) $\psi_\sigma + 2\varphi_{\rm JS} - (6-k)\varpi - k\varpi_5^*$ and $\psi_\sigma + 2\varphi_{\rm JS} - (6-k)\varpi - k\varpi_6^*$ with $k = 0, \ldots, 6$ (the resonances involving ϖ_6^* are not plotted for simplicity, but lie close to the corresponding resonances involving ϖ_5^*). The absence of a visible Nekhoroshev structure in Fig. 11.15 indicates that these resonances should overlap each other. If the period of the great inequality angle $\varphi_{\rm JS}$ is decreased by a factor of 2, this complex of resonances is replaced by that with critical angles $\psi_\sigma + \varphi_{\rm JS} - (3-k)\varpi - k\varpi_5^*$ and $\psi_\sigma - \varphi_{\rm JS} - (3-k)\varpi - k\varpi_6^*$. These resonances have larger coefficients, being of lower eccentricity order, and therefore generate stronger chaos.

It is generally accepted that, in the primordial epochs of the Solar System, Jupiter and Saturn were somewhat closer than at the present time (Fernandez and Ip, 1984). In this case, as pointed out by Nesvorný and Ferraz-Mello (1997b), the period of the great inequality angle would have been shorter. It is then plausible that the primordial asteroids in the A and B regions have been removed during a phase when the period of $\varphi_{\rm JS}$ was half the present one.

In reality, the 2/1 resonance is not completely depleted of asteroids. There are a few asteroids with strongly chaotic orbits, such as 1362 Griqua, which have probably been temporarily captured into resonance from the populations of near-Earth asteroids or of Jupiter family comets. In addition, modern observations are discovering an increasingly large number of small asteroids in the B region at low inclination. 3789 Zhongguo was the first of these asteroids to be discovered. The small size of these objects (not exceeding 10 km) makes their collisional lifetime shorter than the age of the Solar System, so that they should not be primordial asteroids. As shown in the left panel of Fig. 9.19, the Themis family should have injected, at the time of its collisional formation, several asteroids into the 2/1 resonance; on the other hand, in the correct set of coordinates, Zhongguo is perfectly aligned with the Themis family. This allows one to make the appealing conjecture that the asteroids in the B region are all members of the Themis family, injected into resonance in more recent times, after the depletion of the primordial resonant population. This scenario is detailed in Moons *et al.* (1998). Although the statistics are still poor, it seems that the asteroids in the B region have a size distribution that is similar to that of the Themis family and much steeper than that of nonfamily main belt asteroids (Roig, private communication). If confirmed, this would validate

the scenario of the genetic relation of these asteroids with the Themis family.

11.3.3 The 3/2 resonance

The 3/2 resonance is the only low-order mean motion resonance of the main belt that hosts a large population of asteroids (see Fig. 9.12). From the name of the first discovered asteroid, the 3/2 resonant bodies are usually called *the Hildas.*

The existence of the Hildas shows that *resonance* is not synonymous with *gap in the asteroid distribution.* In reality, the 3/2 resonance corresponds to a peak in the asteroid distribution with respect to the background nonresonant population. This is because nonresonant asteroids with $a \sim 4$ AU and non-negligible eccentricity closely approach Jupiter and become rapidly unstable, while resonant bodies are phase-protected from close encounters with the giant planet, as explained in section 9.1.1.

The secular dynamics of the 3/2 resonance is very similar to that of the 2/1 resonance. Figure 11.16 shows the location and the separatrices of the ν_5 and ν_6 resonances. The libration of $\varpi - \varpi_5^*$ is around 180 degrees. The two secular resonances overlap, thus generating a chaotic region that extends over the domain approximately bounded by the lower separatrix of the ν_5 resonance and the upper separatrix of the ν_6 resonance. A region of regular motion is therefore left at moderate eccentricity, where the Hilda asteroids are located. In Fig. 11.16 a few Hildas appear inside the ν_5 resonance, but this is an artifact due to the fact that the separatrices of the ν_5 resonances are plotted for $\varpi - \varpi_5^* = 180°$, while the current osculating elements are used for the asteroids, independently of the value of $\varpi - \varpi_5^*$. If the asteroids were integrated until $\varpi - \varpi_5^* = 180°$, they would all appear on the (a, e) plane outside of the ν_5 resonance.

As Fig. 11.17 shows, also the Kozai and the ν_{16} resonances are present inside the 3/2 commensurability. Figure 11.9 shows on the (e, i) plane at the center of the 3/2 commensurability the leftmost separatrix of the Kozai resonance as well as the location and the borders of the ν_{16} resonance. As one sees, the Hildas also avoid these resonances. Low-order secondary resonances are relegated to the region with $e < 0.15$, and they do not generate an important chaotic region as they do not overlap each other (Michtchenko and Ferraz-Mello, 1995).

Nesvorný and Ferraz-Mello (1997b) have done the same kind of numerical analysis of the 3/2 resonance dynamics that they successfully used for the 2/1 resonance. Their result is shown in the right panel of Fig. 11.15. In this case, the correspondence between the strongly chaotic region shown in

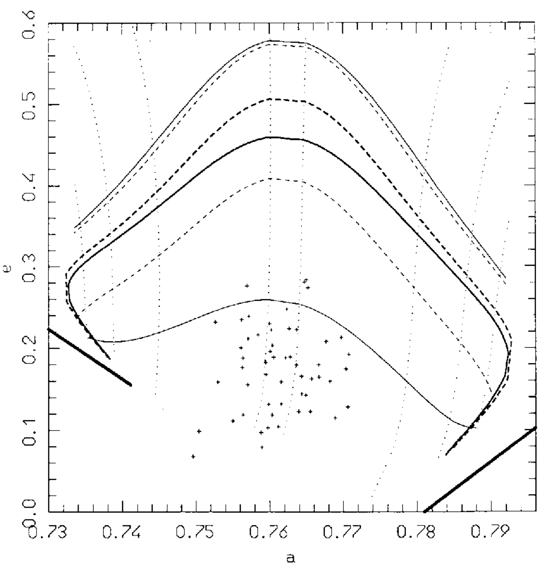


Figure 11.16: The same as Fig. 11.4, but for the 3/2 resonance with Jupiter. The dots denote the osculating elements of the known Hilda asteroids. Reprinted from Fig. 10 of Morbidelli and Moons (1993), with permission from Academic Press.

Fig. 11.15 and the location of the separatrices of the secular resonance plotted in Fig. 11.16 is not good, because the latter are computed for $\varpi - \varpi_5^* = \pi$, while the initial conditions of the test particles integrated by Nesvorný and Ferraz-Mello are taken with $\varpi - \varpi_J = 0$. As a consequence of this choice of the initial phases, the chaotic region generated by the secular resonance overlap is shifted to larger eccentricity in Fig. 11.15.

Figure 11.15 shows that the region that is not affected by secular resonances in the 2/3 commensurability is characterized by much more regular dynamics than the A and B regions in the 2/1 commensurability. In the range $0.2 < e < 0.35$ (at $\varpi - \varpi_J = 0$) at the center of the 2/3 resonance, no trace of chaos is detected, suggesting that most of this region is filled by invariant tori. By integrating all the asteroids until they have $\sigma = \varpi - \varpi_J = 0$ Nesvorný and Ferraz-Mello showed that the vast majority of the Hildas are associated with this stable and regular region.

Figure 11.17 shows the location of the "three body resonances" with critical angles $\psi_\sigma + 3\varphi_{JS} - 9\varpi$ and $2\psi_\sigma + 3\varphi_{JS} - 9\varpi$ in the plane $i = 0$. As for the 2/1 resonance, the resonances with arguments $\psi_\sigma + 3\varphi_{JS} - (9-k)\varpi - k\varpi_5^*$,

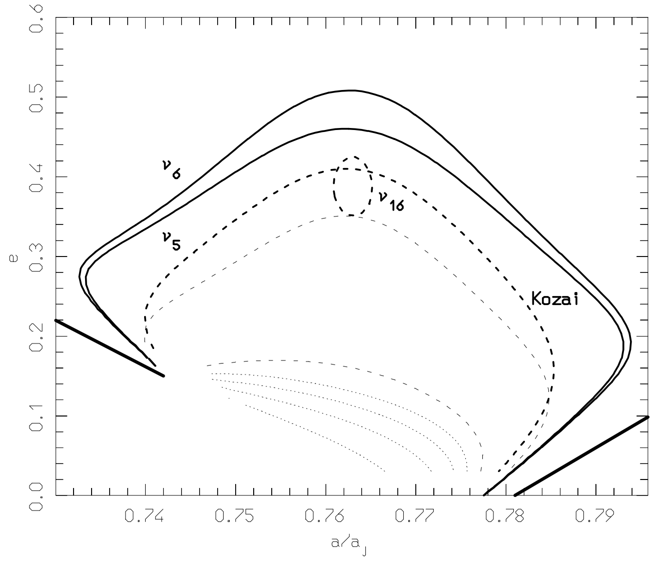


Figure 11.17: The location of the main resonances inside the 3/2 commensurability. The meaning of the bold solid and dashed curves and of the light dotted curves are like in Fig. 11.10. Instead, the light dashed curves denote the location of the three-body resonances $\dot{\psi}_\sigma + 3\dot{\varphi}_{\rm JS} - 9\dot{\varpi} = 0$ and $2\dot{\psi}_\sigma + 3\dot{\varphi}_{\rm JS} - 9\dot{\varpi}$.

$\psi_\sigma + 3\varphi_{\rm JS} - (9-k)\varpi - k\varpi_6^*$ ($k > 0$) and $2\psi_\sigma + 3\varphi_{\rm JS} - (9-k)\varpi - k\varpi_5^*$, $2\psi_\sigma + 3\varphi_{\rm JS} - (9-k)\varpi - k\varpi_6^*$ ($k < 0$) fall in the region between the two resonances reported in Fig. 11.17. There are two reasons for which the three-body resonances do not introduce chaos into the central region of the 3/2 resonance, while they do so in the 2/1 resonance: first, the lowest-order resonance with Saturn that may interfere with the 3/2 resonance with Jupiter is the 15/4 (order 11 in e); second, the period of libration in the 3/2 resonance is shorter than in the 2/1 resonance, which allows only higher-order resonances with the period of the great inequality angle. For these reasons the harmonics related to three-body resonances should have much smaller coefficients than in the case of the 2/1 resonance with Jupiter, so that the individual resonances presumably do not quite overlap each other.

11.4 The major resonances in the Kuiper belt

The study of the dynamics inside the mean motion resonances in the Kuiper belt started only a few years ago, with the discovery of the first resonant objects. It will presumably continue and achieve a more detailed understanding

in the near future, in parallel with the gathering of more observational data on the orbital distribution of the Kuiper belt population. In the following we briefly discuss the present knowledge of the dynamics inside the 2/3 and 1/2 exterior mean motion resonances with Neptune, which seem to be those of major importance for the Kuiper belt.

11.4.1 The 2/3 resonance

About 30% of the objects discovered so far in the Kuiper belt are in the 2/3 resonance with Neptune ($a \sim 39.5$ AU). Taking into account the observational biases that favor the discovery of these objects, Jewitt *et al.* (1998) estimate that the population of bodies in the 2/3 resonance should amount to 10–15% of the total Kuiper belt population in the 30–50 AU range. Pluto, the largest Kuiper belt object known at the present time, is also in the 2/3 resonance. For this reason, the bodies in the 2/3 resonance are usually called *Plutinos*.

The most detailed study of the dynamics inside the 2/3 resonance has been recently done by Nesvorný and Roig (2000). Figure 11.18 shows the location of the main secular, secondary and three-body resonances computed in the (a, e) plane at $i = 0°$ and $\sigma = 180°$ (the stable center of libration in the 2/3 resonance) following the recipe of Section 11.2. In addition, the figure also shows, with a color scale, the logarithm of the maximum Lyapunov exponent computed for a set of test particles initially placed on a regular grid on the (a, e) plane, also at $i = 0°$ and $\sigma = 180°$. Remember from Section 5.2 that the value of the Lyapunov exponent is a direct measure of the strength of chaos. As one sees from Fig. 11.18, the value of the Lyapunov exponent is very well correlated with the location of the various resonances. It is very large close to the borders of the 2/3 resonance, due to the presence of the ν_8 secular resonance and to its overlap with the ν_{18} resonance. It is smaller, but still definitely positive, at small eccentricity ($e < 0.05$), because of the presence of the ν_{18} resonance and of the secondary resonances, as well as along the three-body resonance that corresponds to the 4:1 ratio between the libration period in the 2/3 commensurability and the period of the *great inequality angle* $\varphi_{\rm UN} = 2\lambda_{\rm N} - \lambda_{\rm U}$ of the Uranus–Neptune system. A body in the 2/3 resonance with Neptune is close to the 1/3 resonance with Uranus, so that the harmonics related to three-body resonances may be generated with quite large coefficients by the direct perturbation of Uranus on the body. The 4:1 three-body resonance is in fact a complex of two resonances with critical angles $4\psi_\sigma - \varphi_{\rm UN} + (1-k)\varpi + k\varpi_8^*$, with $k = 0, 1$. These two resonances are both very close to the curve plotted in Fig. 11.18, so that the chaos generated by the 4:1 three-body resonance is very localized in the (a, e) plane, unlike what happens inside the 2/1 mean motion resonance with Jupiter.

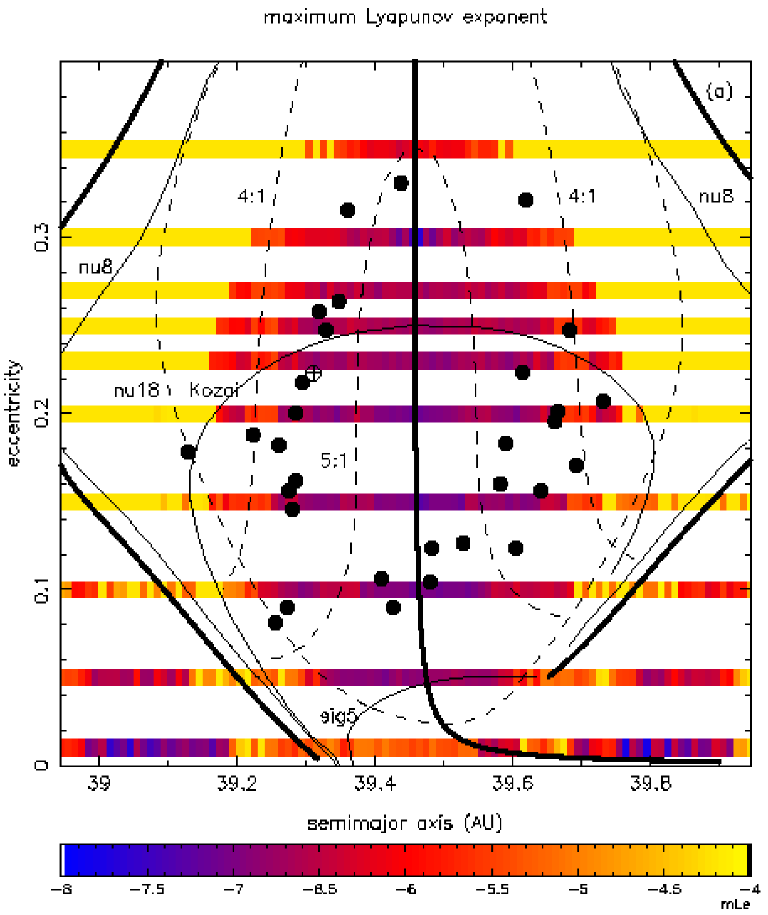


Figure 11.18: Lyapunov exponents and the location of the main resonances inside the 2/3 commensurability with Neptune. The bold curves on both sides of the figure denote the separatrices of the 2/3 resonance. The curves labelled “nu8”, “nu18” and “Kozai” refer to the ν_8, ν_{18} and Kozai resonances respectively; the dashed curves labelled “4:1” and “5:1” denote the location of the three-body resonances $4\dot{\psi}_\sigma - \dot{\varphi}_{\rm UN} \sim 0$ and $5\dot{\psi}_\sigma - \dot{\varphi}_{\rm UN} \sim 0$ (remember that $\dot{\psi}_\sigma > 0$ for exterior resonances); the curve labelled by “sig5” denotes the upper bound of the region affected by secondary resonances of eccentricity order not larger than 5. The filled dots show the semimajor axis and eccentricity attained by the currently known Plutinos when $\sigma = 180°$ and $\omega = 90°$. The crossed circle is the same for Pluto. The colors code the value of the decimal logarithm of the maximum Lyapunov exponent, computed for test particles initially placed on a regular grid in the plane $i = 0°$. Reprinted from Fig. 2 of Nesvorný and Roig (2000), with permission from Academic Press.

At the center of the 2/3 resonance the Lyapunov exponent is of order 10^{-7} y^{-1}. However, this is not the real value of the Lyapunov exponent, but rather the one dictated by the 100 My timespan used for the integrations; on a longer integration timescale, the value of the Lyapunov exponent would presumably further decrease. Thus, the central region of the 2/3 resonance appears characterized by a basically regular dynamics, similarly to the case of the 3/2 resonance with Jupiter (curious coincidence!). Morbidelli (1997) showed with numerical integration of the full equations of motion that the bodies in this region are regular and stable for the age of the Solar System. On the contrary, the bodies that are strongly chaotic at large amplitude of libration escape from the resonance and encounter Neptune in less than 1 Gy. The bodies at moderate amplitude of libration, close to the 4:1 three-body resonance, exhibit slow chaotic diffusion, which may drive them to the strongly chaotic region at large amplitude of libration on a timescale comparable to the age of the Solar System. Nesvorný and Roig (2000) have developed an interesting seminumerical model of this diffusion process.

The position of the Plutinos, projected on the (a, e) plane after having been numerically integrated until the condition $\sigma = 180°$ is satisfied, is shown in Fig. 11.18 and unequivocably indicates their association with the stable central region of the 2/3 resonance. Remember however that the orbital determination of the Kuiper belt objects is still quite inaccurate.

Outside of the reference plane $(i > 0)$ the most important feature of the secular dynamics inside the 2/3 commensurability is the presence of the Kozai resonance. Figure 11.19 shows its separatrices on the (e, i) plane, computed at $\omega = \pm 90°$ at the center of the 2/3 commensurability. The corresponding position of Pluto and of the Plutinos is also plotted. Pluto (the crossed circle in Fig. 11.19) librates inside the Kozai resonance, as first shown by Williams and Benson (1971). Curiously, very few other Plutinos share the Kozai resonance with Pluto, which might be due to the destabilizing effects of low-velocity encounters with the latter (Nesvorný *et al.*, 2000). Pluto has a chaotic motion, as first shown by Sussman and Wisdom (1988). This is due to a resonance between the libration period of ω and the circulation period of $\Omega - \Omega_8^*$, as first shown with numerical integration by Milani *et al.* (1989). This resonance, however, gives rise only to local chaos, so that Pluto is stable (in the sense that it does not leave the Kozai and the 2/3 resonances) for the age of the Solar System (Kinoshita and Nakai, 1996). Figure 11.19 also shows the Lyapunov exponent of test particles with small libration amplitude in the 2/3 resonance, chosen on a regular grid of the (e, i) plane. The only remarkable feature is the presence of a chaotic region at low eccentricity for all inclinations, which is due to the presence of the ν_{18} and of secondary resonances.

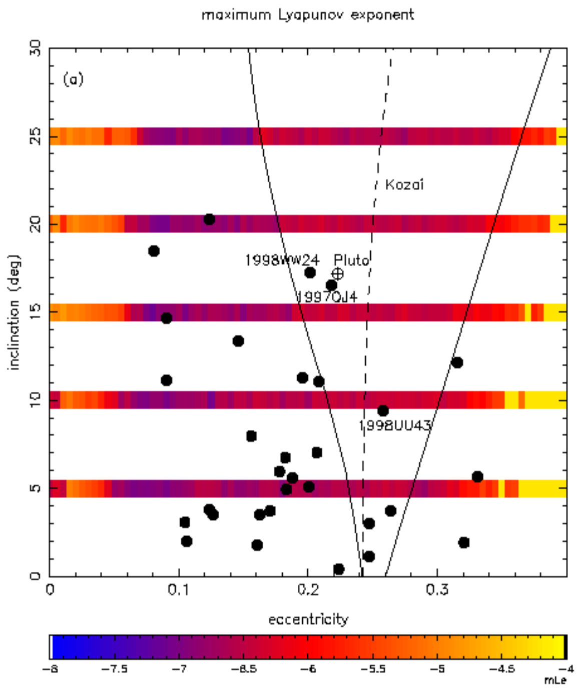


Figure 11.19: The Kozai resonance in the 2/3 commensurability. The bold curves and the dashed curve denote the two separatrices and the center of the Kozai resonance, respectively. The filled dots show the eccentricity and the inclination attained by the currently known Plutinos when $\sigma = 180°$ and $\omega = 90°$. The crossed circle is the same for Pluto. The colors code the value of the decimal logarithm of the maximum Lyapunov exponent, computed for test particles initially placed on a regular grid on the (e, i) plane, at the center of the 2/3 resonance. Reprinted from Fig. 8 of Nesvorný and Roig (2000), with permission from Academic Press.

11.4.2 The 1/2 resonance

The 1/2 resonance with Neptune at 47.8 AU is at the outer border of the presently known population of Kuiper belt objects.

This resonance is peculiar because, as discussed in Section 9.1, it presents two islands of asymmetric libration of σ inside a larger island of symmetric libration (see Fig. 9.4).

The first investigation of the dynamics inside the 1/2 resonance was done by Nesvorný and Roig (2001). To represent both the regions of symmetric and asymmetric libration on the same plot, they computed, for every value

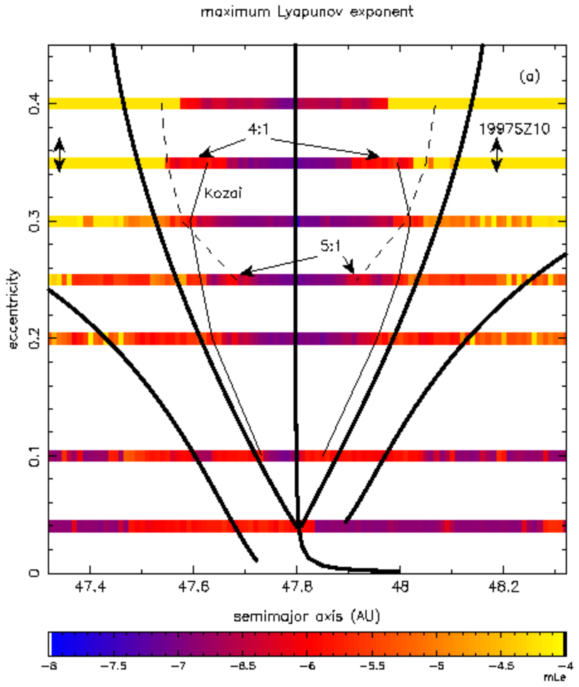


Figure 11.20: The same as Fig. 11.18, but for the 1/2 resonance with Neptune. The leftmost and rightmost bold curves denote the separatrices of the symmetric libration island, while the two bold curves closer to the center mark the position of the separatrices of the islands of asymmetric libration. Reprinted from Fig. 5 of Nesvorný and Roig (2001), with permission from Academic Press.

of the constant action N, the value $\sigma_c(N)$ of the stable equilibrium point of (11.5), at the center of one of the two islands of asymmetric libration; then the correspondence between the values of the actions S, N and those of a, e was computed for $\sigma = \sigma_c(N)$. In fact, it is evident from Fig. 9.4 that both symmetric and asymmetric librating orbits must cross twice the axis $\sigma = \sigma_c(N)$ (or equivalently the axis $\sigma = 2\pi - \sigma_c(N)$).

In this representation, the leftmost and rightmost bold curves in Fig. 11.20 denote the separatrices of the symmetric libration island, while the two bold curves closer to the center of the resonance mark the position of the separatrices of the islands of asymmetric libration. Inside the latter, Nesvorný and Roig have analytically located the position of the Kozai resonance and of the three-body resonance that corresponds to the 5:1 ratio between the period of asymmetric libration and the period of the *great inequality angle*

$\varphi_{\rm UN} = 2\lambda_{\rm N} - \lambda_{\rm U}$. Also shown in Fig. 11.20 are the values of the Lyapunov exponent for test particles initially placed on a regular grid of the (a, e) plane, with $i = 0$ and $\sigma = \sigma_c(N)$. As one sees, the Lyapunov exponent is large for almost all orbits in symmetric libration, revealing strong chaos whose origin has not yet been investigated in detail. Conversely, inside the asymmetric libration island the Lyapunov exponent is in general very small. This allows us to expect, in analogy with the 2/3 resonance, that bodies at the center of the asymmetric libration islands could survive for the entire age of the Solar System. At present, the only candidate body to be in the 1/2 resonance is 1997SZ$_{10}$ but, as shown in Fig. 11.20, it seems to be in a strongly chaotic symmetric libration state. Remember however the orbital indetermination of Kuiper belt bodies is still quite large. Therefore, it is not yet possible to conclude whether the absence of bodies in the stable asymmetric libration islands is an artifact of our limited knowledge of the Kuiper belt population, or a real puzzling feature.

11.5 The 1/1 resonances

The secular dynamics inside a 1/1 resonance with a planet can be studied with the same approach illustrated throughout this chapter. In fact, once the variables (9.9) are introduced, the integrable approximation of the full Hamiltonian is formally identical to (11.5), with $k_{\bar{\jmath}} - k = 1$. Therefore, action–angle variables of type (11.6) can be introduced in the tadpole region and in the horseshoe region, and the study of the secular dynamics then follows as discussed in Section 11.2.

As an example of its application, Fig. 11.21 shows the dynamics related to the ν_6 resonance at the center of the tadpole region in the 1/1 resonance with Saturn. The "real" ν_6 resonance – that is, characterized by stable and unstable equilibrium points and a separatrix surrounding the libration region – occurs at larger eccentricity than shown in these diagrams. But at low eccentricity a stable equilibrium point appears at $\varpi - \varpi_6^* = 60°$. The eccentricity of this equilibrium point is close to the value of the $M_{6,6}$ coefficient in (7.10). There is no separatrix at the boundary between the libration and the circulation regions of $\varpi - \varpi_6^*$. In a polar representation, the secular dynamics is simply made of quasi-circular trajectories, offset with respect to the center $e = 0$.

As a consequence of the existence of this equilibrium point, all orbits with initial $e \sim 0$ reach an eccentricity larger than 0.15 when $\varpi - \varpi_6^* = 60°$. Remember from Section 9.2.2 that at the center of the tadpole region the dynamics is chaotic above 0.12, for the overlap with the 2/5 mean motion resonance with Jupiter. Chaos acts on a timescale that is shorter than the

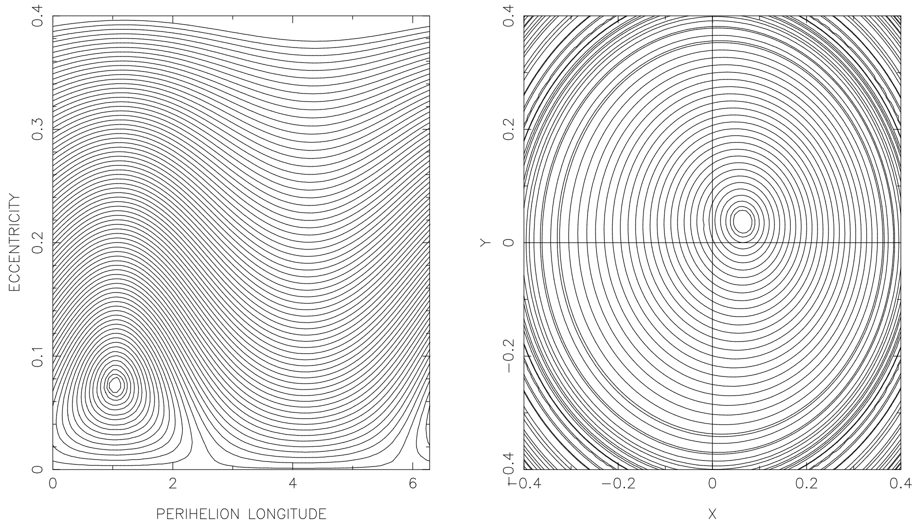


Figure 11.21: The ν_6 secular resonance dynamics at the center of the tadpole region in the 1/1 resonance with Saturn. The left panel shows a Cartesian representation of the dynamics, while in the right panel a polar representation is used (that is $x = e\cos\varpi, y = e\sin\varpi$). Here ϖ is expresed relative to ϖ_6^*. The direct perturbations exerted by Jupiter, Uranus and Neptune – double averaged over the mean longitudes of the planet and of the small body – have also been included in this computation. Courtesy of D. Nesvorný.

precession period of $\varpi - \varpi_6^*$, because it is generated by mean motion degrees of freedom, and therefore it allows the body to escape from the 1/1 resonance before that the secular precession may bring the eccentricity back to smaller values.

At larger amplitude of libration in the tadpoles region, the chaotic region generated by the overlap with the 2/5 resonance with Jupiter is confined to larger eccentricities (see Fig. 9.14). As a consequence, the secular dynamics is not powerful enough to drive objects initially at $e \sim 0$ into the chaotic region.

In conclusion, the combination of the secular dynamics with the effects generated by the overlap with the 2/5 resonance with Jupiter does explain the absence of Saturn's Trojans with small amplitude of libration, but cannot explain that of the bodies with larger libration amplitude. Long-term full numerical simulations done by Holman and Wisdom (1993) show that large amplitude tadpole librators are indeed stable on a several 2×10^7 y timescale,

while low libration amplitude tadpoles are rapidly eliminated. Therefore, the cause of the absence of objects on large amplitude tadpole orbits should be sought in the (still largely unknown) mechanisms that sculpted the early Solar System.

Detailed numerical maps of the stability regions in the 1/1 resonances with Saturn, Uranus and Neptune are provided in Nesvorný (2001).

Chapter 12

GLOBAL DYNAMICAL STRUCTURE OF THE BELTS OF SMALL BODIES

12.1 Detection of the chaotic zones

In the previous chapters we have analyzed the dynamical structure of two-body and three-body mean motion resonances of various orders in the eccentricity. In this final chapter we study how all these resonances are nested in the asteroid belt and in the Kuiper belt and the dynamical structure that results from their interplay. In principle one could analytically compute the separatrices of each resonance (or of each component of each resonant multiplet) in order to obtain a picture similar to Fig. 9.12, but extended to two-body resonances of much higher eccentricity order and to three-body resonances. In practice this is infeasable, as one should compute and study thousands of mean motion resonant normal forms.

The easiest and most efficient way to detect the resonant structure of the belts of small bodies is to make numerical explorations, using the tools detailed in Chapter 7. The knowledge of resonant dynamics acquired in the previous chapters will allow us to interpret the results and understand their implications.

Figure 12.1 is the analog of Fig. 10.8, but extends the exploration to the entire asteroid belt. 5700 test particles were initially set on a regular grid in the semimajor axis that ranges from 2.1 to 3.24 AU with a 2×10^{-4} AU spacing. The initial eccentricity was chosen equal to 0.1, while the initial inclination and phase angles were all set equal to zero with respect to the ecliptic and the vernal point. The test particles were integrated over 2.3 My, together

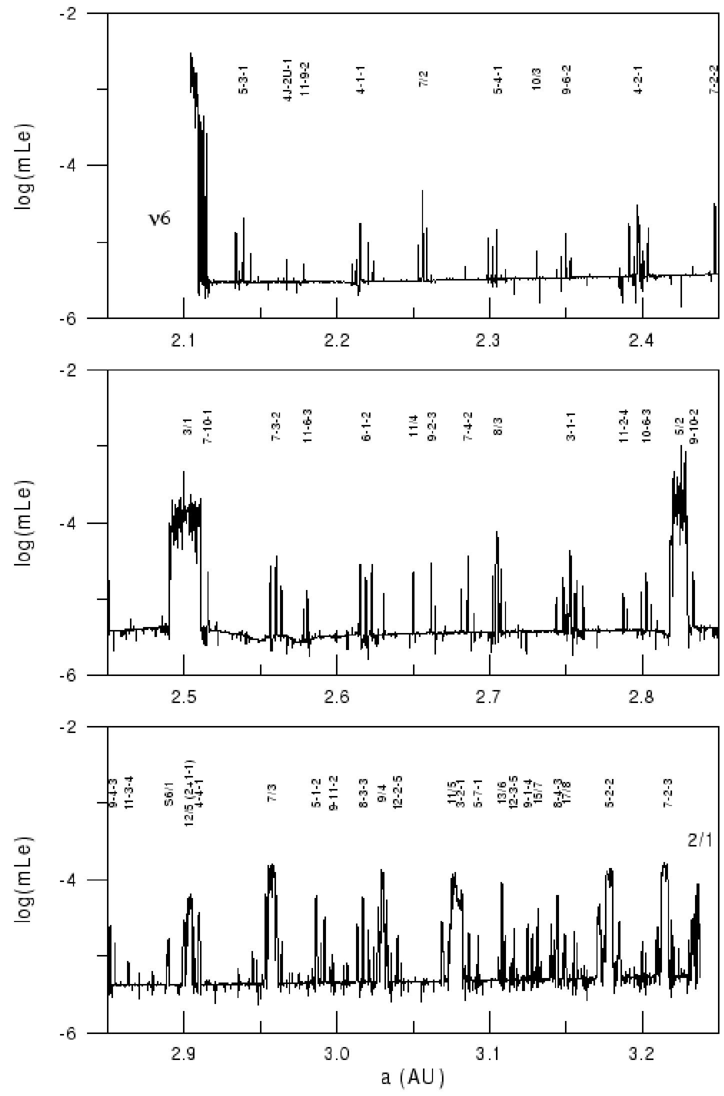


Figure 12.1: Maximum Lyapunov exponent as a function of the semimajor axis for asteroids with initial eccentricity equal to 0.1 and initial inclination equal to 0. Only the perturbations provided by the four giant planets are taken into account in the model. The unit of the MLE is y^{-1}. Labels m/n denote the m/n two-body resonances with Jupiter, while labels $m\ n\ k$ denote the three-body Jupiter–Saturn–Asteroid resonances that correspond to the equality $m\dot{\lambda}_{J} + n\dot{\lambda}_{S} + k\dot{\lambda} = 0$. Moreover the label S6/1 denotes the 6/1 resonance with Saturn and 4J-2U-1 refers to the three-body Jupiter–Uranus–asteroid resonance. Reprinted from Fig. 1 of Morbidelli and Nesvorný (1999), with permission from Academic Press.

with their variational equations in order to estimate the maximum Lyapunov exponent, under the gravitational influence of the four giant planets. The figure shows the resulting estimate of the Lyapunov exponent (in logarithmic scale) as a function of the initial semimajor axis. As usual, the figure shows several peaks which rise from a background level; remember from Section 5.2 that peaks, dips and discontinuities correspond to chaotic or resonant regions, while a flat background level reveals a "regular" region, and its value ($\sim 10^{-5.3}$ y^{-1} in Fig. 12.1) is dictated by the limited integration timespan: increasing the latter, the background level would generally decrease.

Figure 12.1 shows that the chaotic regions become denser with increasing semimajor axis. This is due to two reasons. First, the location of mean motion resonances of a given order become denser approaching Jupiter. Second, resonances of higher eccentricity order can be detected, thanks to the fact that the Lyapunov exponent is correlated with the size of the coefficients of the resonant harmonics (see Section 12.3), and the latter grows with decreasing distance from the main perturber.

Note also that the peaks are located according to a peculiar structure: they form multiplets with a leading component at the center and secondary components on each side. This structure should not be confused with the multiplet structure of high-order mean motion resonances (see Section 9.3), so that we prefer to refer to it as a *supermultiplet* structure. The latter is a consequence of Jupiter and Saturn being close to the 5/2 mean motion resonance. The combination $\varphi_{\mathrm{JS}} = 5\lambda_{\mathrm{S}} - 2\lambda_{\mathrm{J}}$ is an angle which circulates with positive derivative with a period of about 880 y. This period is much longer than the typical asteroid orbital period, so that, for any fixed integer numbers m, m_{J} and m_{S}, the resonances given by the relations $m\dot{\lambda} + (m_{\mathrm{J}} + 2k)\dot{\lambda}_{\mathrm{J}} + (m_{\mathrm{S}} - 5k)\dot{\lambda}_{\mathrm{S}} \sim 0$ with different integer k, must be located close to each other (Murray *et al.*, 1998). The leading component of each supermultiplet is most often associated with the resonance of minimal eccentricity order or with the resonance with $m_{\mathrm{S}} - 5k = 0$. In Fig. 12.1, labels show the integers $m\ m_{\mathrm{J}}\ m_{\mathrm{S}}$ of the central components of most supermultiplets. Starting from a central component, the resonances with $k = 1, 2, \ldots$ (resp. $k = -1, -2, \ldots$) are situated on the right (resp. left) side of the central component if m is negative, and on the left (resp. right) side if m is positive.

Figure 12.1 shows that most of the chaotic regions are associated with either ordinary two-body mean motion resonances with Jupiter ($m_{\mathrm{S}} = 0$) or three-body Jupiter–Saturn–asteroid resonances, although a few smaller peaks can be associated with three-body Jupiter–Uranus–asteroid resonances of low order. Conversely, no peak is apparently associated with a secular resonance. Secular resonances, in fact, should give Lyapunov exponents no larger than $\sim$

10^{-6} y^{-1}, because of the slow motion of the critical angles of secular resonances (whose typical periods exceed $\sim$ 1 My). Only the ν_6 secular resonance is visible in Fig. 12.1, at the inner border of the asteroid belt. The test particles in this resonance escape to very large eccentricity (see Section 8.4.2) and their integration is stopped before 2.3 My, which results in an estimate of their Lyapunov exponent exceeding 10^{-4} y^{-1}.

As discussed in Chapters 8 and 9, the widths in the semimajor axis of all mean motion resonances – with the exception of those of order zero in the eccentricity – increase with increasing asteroid eccentricity. As a consequence, the amount of chaos in the belt is enhanced at large eccentricity; the chaotic regions associated with different resonances may overlap making the belt globally chaotic. Beyond 3.1 AU this phenomenon has already been shown in Fig. 10.8. In the inner belt, Morbidelli and Nesvorný (1999) have checked that increasing the initial asteroidal eccentricity to 0.2 just causes the overlapping of the resonances within the same supermultiplet, but distinct supermultiplets are still well separated. Therefore, the inner asteroid belt, in the approximation where only the outer planets are considered, appears to be characterized by chaotic bands, separated by large regular regions.

However, the chaotic structure of the inner belt changes dramatically if the effects of the inner planets are also considered. Figure 12.2a shows the Lyapunov exponents in the inner belt resulting from a 2.3 My integration that includes all of the Solar System planets except Mercury. It can be directly compared with the top panel of Fig. 12.1 because the initial conditions of the integrated test particles are identical. The comparison shows in a striking way that the inner planets create a very large number of chaotic regions, which become denser with decreasing semimajor axis. Moreover, the background level of the Lyapunov exponent among the various peaks appears very irregular, in contrast with the flat shape that it had when only the outer planets were considered. This is an additional indication of a dense concentration of effective resonances and of the global chaoticity of the region. At semimajor axes smaller than 2.16 AU no background level is visible: for all test particles the computation of the Lyapunov exponent reaches a positive limit value within the integration timespan. Notice that the Lyapunov exponent tends to increase with decreasing semimajor axis, until the ν_6 secular resonance is reached at 2.1 AU. The latter is moved to smaller semimajor axis with respect to Fig. 12.1, because the presence of the inner planets slightly speeds up the precession rates of the asteroids' longitudes of perihelia.

The generally positive value of the Lyapunov exponent is *not* due to close encounters between the test particles and Mars. Figure 12.2b shows the minimal approach distance to Mars over the integration time as a function of the

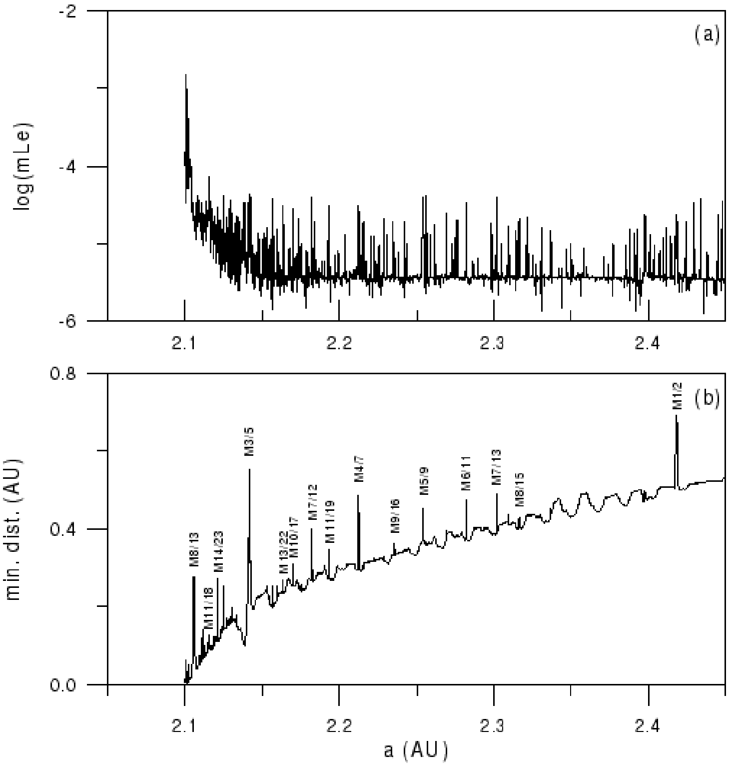


Figure 12.2: Top panel: the same as the top panel of Fig. 12.1, but taking into account also the perturbations exerted by the terrestrial planets. Bottom panel: the test particles' minimal distance to Mars as a function of the initial semimajor axis. The peaks reveal the location of the outer mean motion resonances with Mars. The labels Mn/k denote the n/k outer resonances with Mars. Reprinted from Fig. 3 of Morbidelli and Nesvorný (1999), with permission from Academic Press.

particles' semimajor axis: the minimal distance is larger than 1 Martian Hill's sphere (7.25×10^{-3} AU) for all particles with $a \geq 2.103$ AU and larger than 3 Martian Hill's spheres for $a \geq 2.106$ AU. Therefore, almost none of the test particles undergoes close encounters with Mars, and their positive Lyapunov exponent must be due to resonances with the inner planets. To identify some of these, Fig. 12.2b is also very useful: the peaks that it shows reveal the presence of external mean motion resonances with Mars. Only in these resonances, in fact, does the body avoid the closest approaches to the planet due to the protection mechanisms that they provide (see Section 9.1.1). Figure 12.2b labels the main external mean motion resonances with Mars. Note that the

peaks in Fig. 12.2b are always correlated with peaks in the Lyapunov exponent profile of Fig. 12.2a. Moreover, many small peaks appear in Fig. 12.2b for $a < 2.15$ AU, showing that high-order external mean motion resonances with Mars accumulate in this region. This explains why the Lyapunov exponent is always positive in this semimajor axis range.

However the external mean motion resonances with Mars are not sufficient to explain the entire chaotic structure of the inner belt. In fact, there are many more peaks in the Lyapunov exponent profile than in the minimal approach distance curve. Morbidelli and Nesvorný (1999) have checked that many of these peaks are in fact related to three-body Mars–Jupiter–asteroid mean motion resonances. Conversely, the mean motion resonances with the Earth or Venus and the multibody resonances involving combinations of the orbital frequencies of the terrestrial planets do not seem relevant to the origin of chaos.

Because the mass of Mars is small ($\sim 3.3 \times 10^{-7}$ solar masses) it might be surprising that high-order Martian mean motion resonances or Mars–Jupiter–asteroid three-body resonances are so relevant for asteroid motion. However, one should not forget that the asteroids in the inner belt come close to Mars, so that its small mass is compensated by the small distance. Moreover, because of the smaller approach distance, the perturbation function related to Mars has a radius of analyticity that is much smaller than that of the perturbation function related to Jupiter, so that the coefficients of its Fourier expansion decay much more slowly with the order of the harmonics. Finally, because the eccentricity of Mars is comparable to the eccentricity of the considered asteroids, the coefficients of the harmonics of the different critical angles, which correspond to different relative powers of $e_{\rm M}$ and e (resp. the eccentricity of Mars and of the asteroid), are of the same order of magnitude, producing a strong time modulation of the amplitude of the resonances (see Section 9.4). As an example, the coefficient of the harmonic $\cos(16\lambda_{\rm M} - 27\lambda + 3\varpi_{\rm M} + 8\varpi)$ of the 16/27 mean motion resonance with Mars is $-5.6032 \times 10^7 \varepsilon_{\rm M} e_{\rm M}^3 e^8 / a_{\rm M}$, where $\varepsilon_{\rm M}$ and $a_{\rm M}$ denote the mass and the semimajor axis of Mars. Therefore, when both e and $e_{\rm M}$ are equal to 0.1, the width of the 16/27 resonance is equal to $\sim 1.6 \times 10^{-4}$ AU. For comparison, the 15/4 mean motion resonance with Jupiter, which is located very close to the 16/27 resonance with Mars, has the coefficient of its leading harmonic $\cos(15\lambda_{\rm J} - 4\lambda - 6\varpi_{\rm J} - 5\varpi)$ equal to $-6857\varepsilon_{\rm J} e_{\rm J}^6 e^5 / a_{\rm J}$ which, for $e = 0.1$ and $e_{\rm J} = 0.05$, gives the resonance a width of only 6.5×10^{-6} AU. Note that the test particles used for the computation of the Lyapunov exponent in Figs. 12.1 and 12.2 have an initial separation of 2×10^{-4} AU, so that in principle these figures could miss a large number of peaks related to thinner resonances.

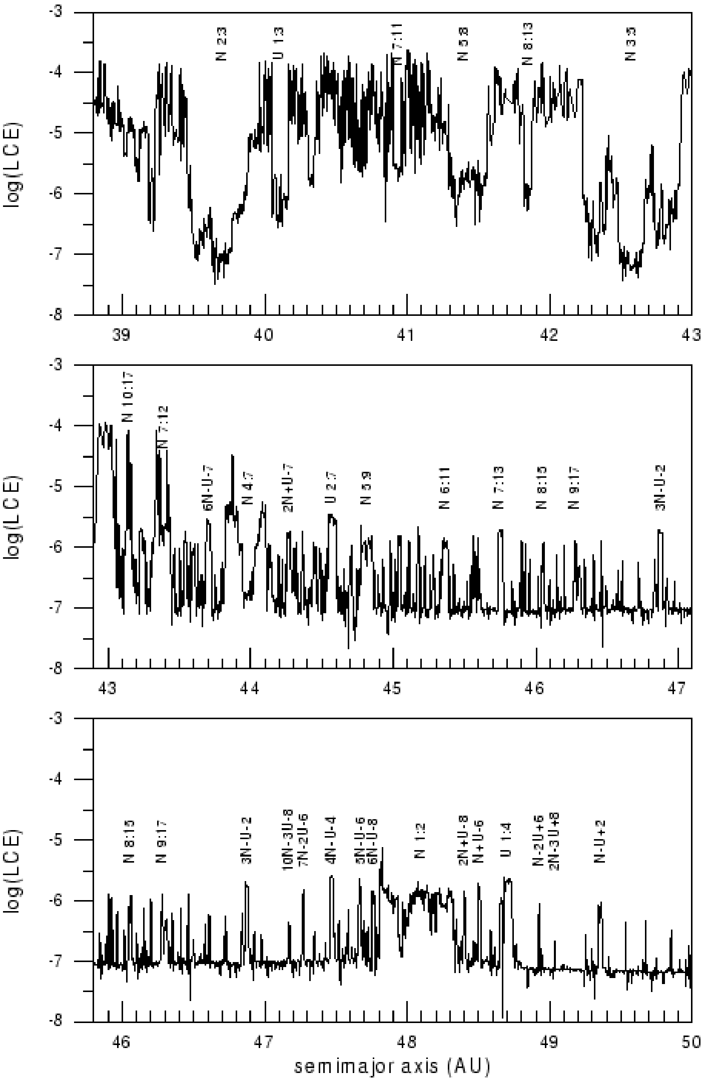


Figure 12.3: The same as Fig. 12.1, but for the Kuiper belt. Labels $Nm : n$ and $Um : n$ denote the m/n two-body resonances with Neptune and with Uranus, respectively. Labels mN+kU+j denote the three-body Neptune–Uranus–body resonances that correspond to the equality $m\dot{\lambda}_{\rm N} + k\dot{\lambda}_{\rm U} + j\dot{\lambda} = 0$. Reprinted from Fig. 2 of Nesvorný and Roig (2001), with permission from Academic Press.

An analogous study has been recently done by Nesvorný and Roig (2001) for the Kuiper belt. They set 2800 test particles regularly spaced in the semimajor axis between 38.8 and 50 AU, with initial eccentricity equal to 0.1 and inclination and phase angles equal to zero, and integrated their evolution under the effect of the four giant planets for 100 My. Figure 12.3 shows the resulting maximum Lyapunov exponent as a function of the initial semimajor axis and reveals that the Kuiper belt also has a complex chaotic structure.

In analogy to Fig. 12.1, most features visible in Fig. 12.3 are related to mean motion resonances, the secular resonances typically having smaller Lyapunov exponents due to their longer dynamical timescales. The only exception is the 40–42 AU interval, where the ν_8, ν_{17} and ν_{18} secular resonances overlap (Knežević *et al.*, 1991; Fig. 8.10) and force many test particles to escape to Neptune-crossing orbit within 100 My.

Some of the mean motion resonances are labeled in Fig. 12.3. While for $a < 46.5$ AU the largest peaks are associated with the mean motion resonances with Neptune, for $a > 46.5$ AU, most peaks correspond to the three-body resonances with Neptune and Uranus. Note that the latter are placed on both sides of the 1/2 resonance with Neptune, following a supermultiplet structure analogous to those discussed above for the asteroid belt. This is a consequence of Uranus and Neptune being close to the mutual 1/2 resonance, so that the angle $\varphi_{\mathrm{UN}} = 2\lambda_{\mathrm{N}} - \lambda_{\mathrm{U}}$ slowly circulates with positive derivative and a period of about 4230 y.

Notice also from Fig. 12.3 that the chaotic regions become denser with decreasing semimajor axis. This is a phenomenon analogous to the accumulation of mean motion resonances with Mars in the inner asteroid belt and of mean motion resonances with Jupiter in the outer asteroid belt. Notice however that the major mean motion resonances with Neptune in the inner Kuiper belt (2/3, 7/11, 5/8, 8/13, 3/5, but also the 1/3 resonance with Uranus) cause a general decrease of the value of the Lyapunov exponent. The orbits in these resonances are less chaotic than the neighboring nonresonant orbits, because of the phase-protection mechanism offered by these resonances and of the different secular dynamics. The dynamics inside the 2/3 and 1/2 resonances has been already described with some detail in Section 11.4.

An exploration of the dynamics of small bodies, extended to basically the entire Solar System, has recently been accomplished by Robutel and Laskar (2000) using frequency analysis. They numerically integrated the evolution of 192,000 test particles, initially placed on the (a, e) plane, with inclination and all phase angles equal to zero. The grid of initial conditions covered 80 values evenly spaced between 0 and 1 in eccentricity, 200 values in each of the semimajor axis intervals [0.38,0.9], [0.9,2.0], [2,5], [5,10], and 1800 values in

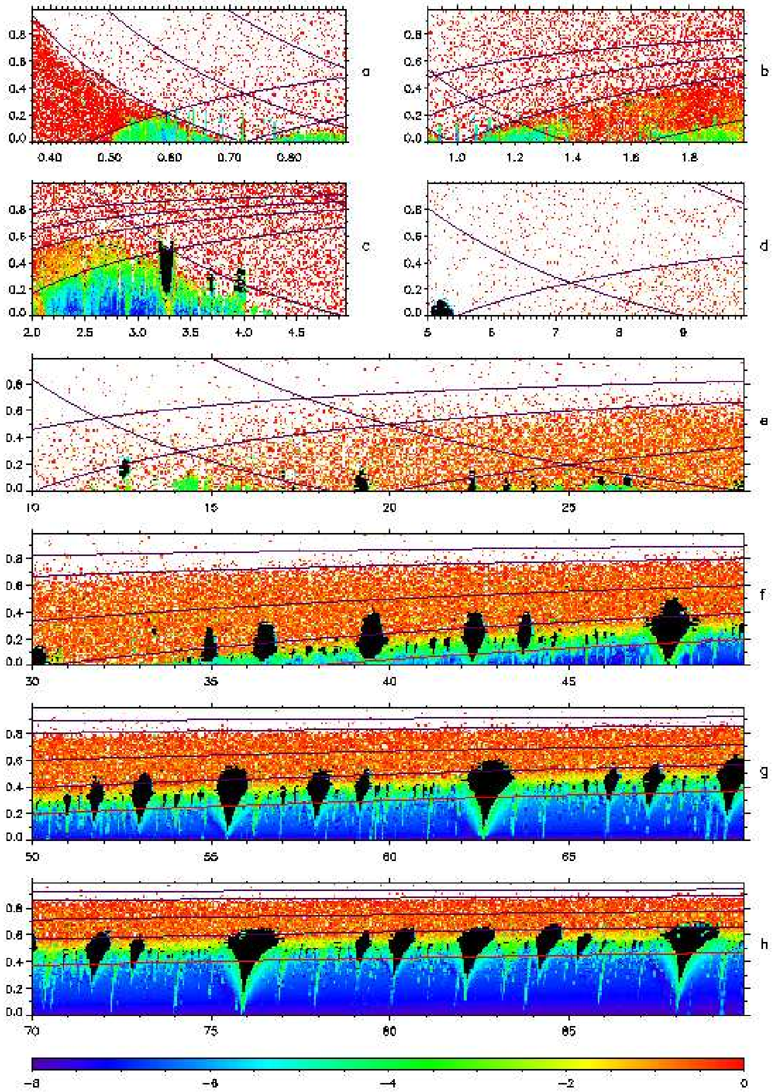


Figure 12.4: Frequency map of small body dynamics over all the Solar System. Initial conditions are chosen with inclination and all phase angles equal to 0. The color codes the value of $\text{Log}\,\sigma$ (unit: My^{-1}), where σ is the relative change of the frequency of the particle's mean longitude. The black color denotes bodies that are recognized as being in some two-body mean motion resonance. The solid curves mark the planet-crossing conditions. Reprinted from Fig. 2 of Robutel and Laskar (2001), with permission from Academic Press.

the semimajor axis range [10,100] (AU). The integrations covered 0.5 My for the test bodies in the inner solar system, under the gravitational influence of all planets, and 2 My for the test bodies in the outer solar system, under the influence only of the giant planets. The integration time [0,T] was divided into two subintervals [0,$T/2$], [$T/2$,T], on each of which the frequency of the mean longitude λ of each small body was numerically determined. Denoting by $n^{(1)}$ and $n^{(2)}$ the frequencies measured on the first and second subinterval respectively, the frequency change is defined as $\sigma = 1 - n^{(2)}/n^{(1)}$. Figure 12.4 shows the value of $\mathrm{Log}\,\sigma$ as a function of the initial conditions, using a color code and the additional convention that the initial condition of a body undergoing a close encounter with a planet within the integration timespan is colored in white (the same color as the background), while the initial conditions of bodies recognized to be in mean motion resonances are plotted in black, whatever their dynamical behavior (regular or chaotic).

Figure 12.4 provides a nice global view of the dynamics of small bodies in the Solar System. The resolution is much lower than that of Figs. 12.1, 12.2 and 12.3, so that the fine structure given by high-order mean motion resonances and three-body resonances is not recognizable. However, the figure provides a useful indication of where regular or quasi-regular orbits are most likely to be located. It is remarkable that in the asteroid belt these orbits (colored in dark blue) can be found in significant number only in the region with $2.15 < a < 3.2$ AU and $e < 0.2$. Concerning the Kuiper belt, only the region with $a > 50$ AU shows an essentially regular structure over a large range of eccentricities. Unfortunately, no real objects have yet been detected in the deep Kuiper belt.

A second remarkable consideration is that the moderately chaotic regions (colored in green in Fig. 12.4) in the asteroid belt and in the Kuiper belt are actually inhabited by real objects, suggesting that chaos is too weak to cause the escape of the majority of objects over the age of the Solar System (see Section 12.2); conversely, the green regions that are visible in Fig. 12.4 in between the terrestrial planets and in between the giant planets are not associated with populations of small bodies. This difference might provide important indications on the process of planetary formation and/or on the primordial evolution of the Solar System.

12.2 Chaotic diffusion and macroscopic instability

If the detection of the chaotic structure can be achieved with relatively short numerical integrations, the study of the effects of weak chaos on the long-term dynamical evolution of the orbits requires numerical integrations on timescales

as long as possible. Numerical integrations covering 100 My or 1 Gy are now possible thanks to the new generation of fast computers and fast integration algorithms (Wisdom and Holman, 1991; Levison and Duncan, 1994; Duncan *et al.*, 1998; Chambers, 1999). We therefore begin to have a first, partially complete, view of the evolution of the asteroid and Kuiper belts on timescales comparable to the age of the Solar System, which will be rapidly improved in the near future.

The slow chaotic diffusion of orbits in the inner asteroid belt has been studied in the papers by Migliorini *et al.* (1998) and Morbidelli and Nesvorný (1999). A sample of 412 real asteroids has been integrated over 100 My. These bodies have been chosen among those with osculating perihelion distance smaller than 1.8 AU, semimajor axis smaller than 2.5 AU, inclination smaller than 15 degrees and which are not Mars-crossers in the first 300,000 y. The last property allows the exclusion from the sample population of the "hidden Mars-crossers", namely those bodies that cross the orbit of Mars on a short timescale due to the secular oscillation of their orbital elements. The perturbations of all planets have been taken into account, with the exception of Mercury.

To highlight small changes of the orbital elements over long timescales, the asteroid's *proper elements* and their change over the integration timespan have been numerically computed. This was done by averaging the orbital elements, originally output every 500 y, using a running window that covers a 10 My timespan. In other words, denoting generically the semimajor axis or the eccentricity by x, the proper value $x_p(t)$ has been computed as

$$x_p(t) = \frac{1}{N} \sum_{t'=t-5My}^{t'=t+5My} x(t') \tag{12.1}$$

where N denotes the number of output values of x over the considered time interval; the time t has been incremented by 10^5 y steps from the initial value $t = 5$ My. A 10 My window is long enough to average out all important quasi-periodic oscillations of the osculating elements, and therefore – as explained in Section 5.4.4 – the change over time of proper elements reveals non-quasi-periodic evolution, i.e. *chaotic diffusion.* Conversely, regular bodies, having only quasi-periodic oscillations of their osculating elements, have proper elements that are constant with time.

The initial proper semimajor axis and eccentricity of the bodies integrated by Migliorini *et al.*, computed using formula (12.1) on the first 10 My of integration, are shown in the top panel of Fig. 12.5. Red denotes bodies that will become Mars-crossers within the integration time. Only three bodies have a dynamical lifetime shorter than 10 My, so that their initial proper elements

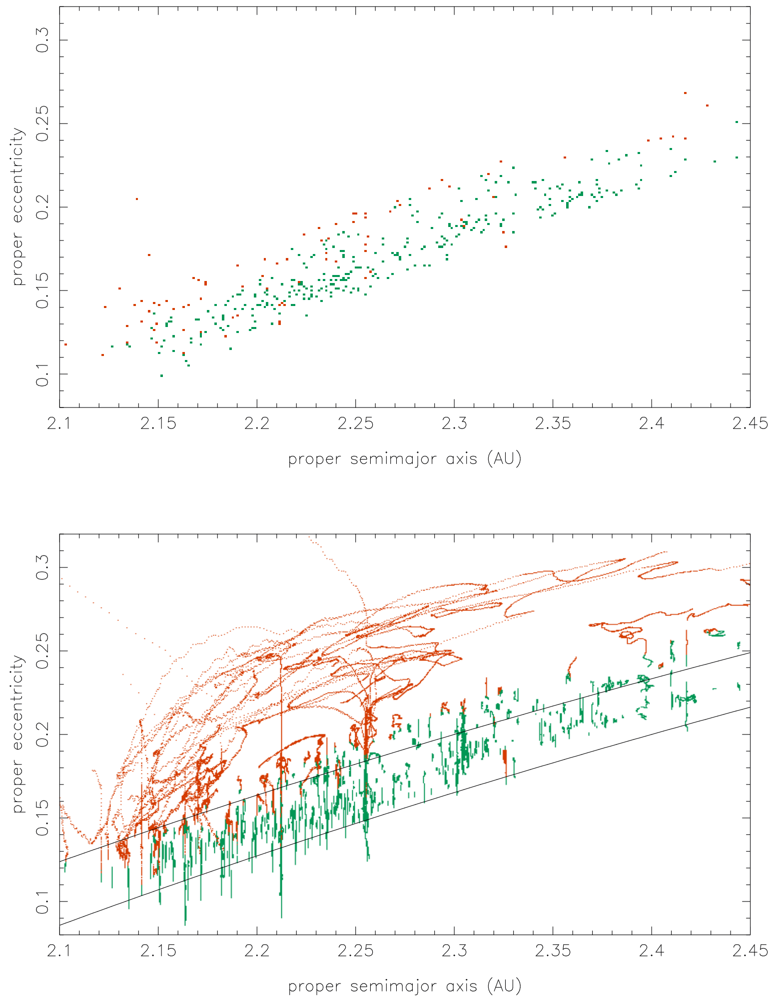


Figure 12.5: Top panel: The proper semimajor axis and eccentricity computed over the first 10 My of evolution of the integrated asteroids in the inner belt. Red dots denote objects that will become Mars-crossing before the end of the simulation. Bottom panel: the subsequent evolution of the integrated bodies in the proper (a, e) plane. Regular bodies appear as a dot, while chaotic bodies migrate, leaving a trace in the plot. The red color distinguishes the evolution of the bodies after that they become Mars-crossers. The two curves denote a proper perihelion distance equal to 1.92 and 1.84 AU. Reprinted from Fig. 8 of Morbidelli and Nesvorný (1999), with permission from Academic Press.

cannot be computed. The bottom panel of Fig. 12.5 shows the evolution of the proper semimajor axis and eccentricity of the integrated bodies: only very few bodies have regular dynamics (those which appear as a dot), while the vast majority exhibits macroscopic diffusion in eccentricity – that is, a significant change of the proper eccentricity. This is in agreement with the result of Fig. 12.2 that most of the bodies in the inner belt are chaotic. In the bottom panel of Fig. 12.5, the red color is used to distinguish the evolution of the bodies when they are in the Mars-crossing regime, namely when their osculating perihelion distance becomes smaller than 1.665 during the timespan covered by the computation of the corresponding proper elements. Under the effect of Martian encounters, these bodies start to random-walk in the semimajor axis, roughly following a curve of invariant Tisserand parameter with respect to Mars (see Öpik 1976). Consequently, their proper semimajor axes change with time. This change is moderate for shallow Mars-crossers, and much bigger for deep Mars-crossers. Conversely, for most bodies chaotic diffusion keeps the proper semimajor axis basically constant until the Mars-crossing status is reached. This implies that the bodies diffuse in eccentricity always staying in the same mean motion resonance or alternating among closely located resonances. Only the few nonMars-crossing asteroids that show macroscopic diffusion of the proper semimajor axis should migrate along a chain of overlapping resonances.

The amount of chaotic diffusion is not the same over all the inner asteroid belt. Figure 12.5 shows regions characterized by large proper eccentricity variations, and a background characterized by smaller, but still nonnegligible, eccentricity changes. To distinguish, we denote hereafter the first as the *main diffusion tracks* and the second as the *diffusion background*. The main diffusion tracks are: (i) at 2.256 AU, related to the almost coincident locations of the 7/2 mean motion resonance with Jupiter and of the 5/9 mean motion resonance with Mars (curiously, the latter can be computed to be more effective than the former); (ii) at 2.213 AU, related to the presence of the 4/7 mean motion resonance with Mars; (iii) the region with semimajor axis smaller than 2.17 AU, where Martian mean motion resonances overlap each other and give rise to global chaos, as shown in Fig. 12.2. The 3/5 mean motion resonance with Mars at 2.142 AU is the most important among these resonances. Moving to the outer part of the inner belt, the two remarkable diffusion tracks at 2.398 and 2.419 AU are respectively related to the $4\ -2\ -1$ three-body Jupiter–Saturn–asteroid resonance and to the 1/2 mean motion resonance with Mars.

Figure 12.6 compares the semimajor axis distribution of the Lyapunov exponent with the semimajor axis distribution of the proper eccentricity change.

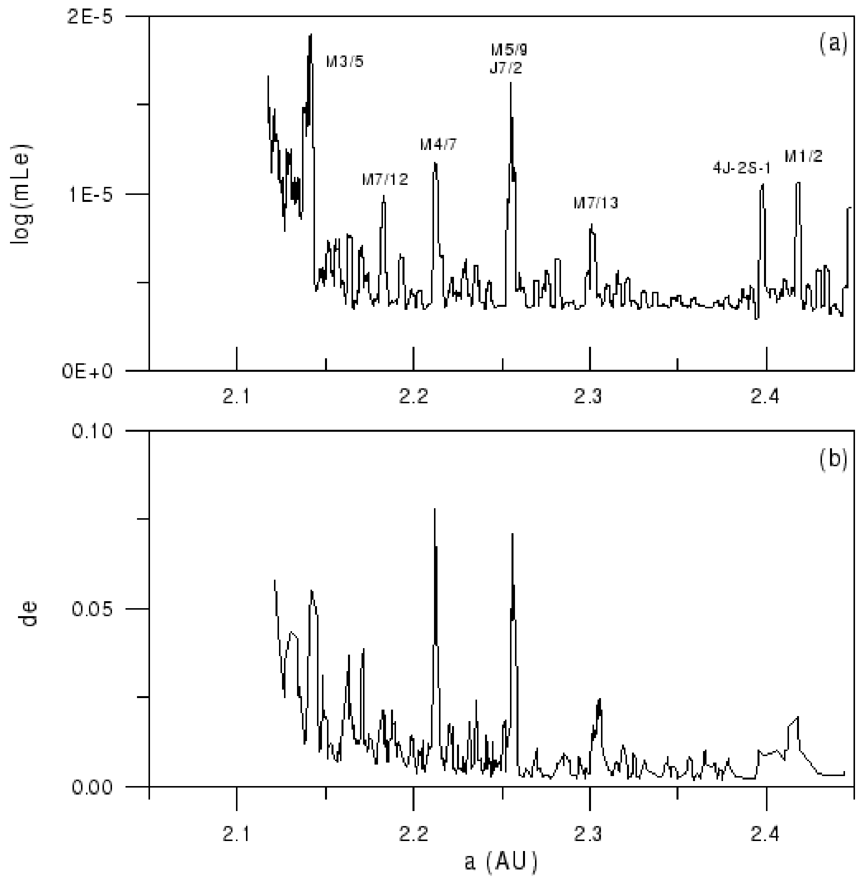


Figure 12.6: Comparison between the semimajor axis distribution of the maximum Lyapunov exponent (top panel) and of the proper eccentricity change (bottom panel). Labels Mn/m and Jm/n denote the m/n mean motion resonances with Mars and Jupiter respectively, while 4J−2S−1 denotes the Jupiter–Saturn–asteroid three-body resonance $4\dot{\lambda}_J - 2\dot{\lambda}_S - \dot{\lambda} \sim 0$. Reprinted from Fig. 9 of Morbidelli and Nesvorný (1999), with permission from Academic Press.

The latter is obtained by computing, for each test particle of Fig. 12.5, the maximal change of proper eccentricity de that occurs *before* being a Mars-crosser. Both distributions in Fig. 12.6 have been smoothed using a suitable running window average. The comparison between Fig. 12.6a and Fig. 12.6b is striking. Not only do the main peaks – associated with the main diffusion tracks discussed above – correspond, but also the minor peaks that structure

the diffusion background appear to correspond to peaks in the Lyapunov exponent profile. In particular, in both panels peaks are particularly dense in the region with semimajor axes smaller than 2.25 AU, while they are less prominent and more separated in the more regular region between 2.3 and 2.4 AU. Moreover, both the MLE and de tend to increase with decreasing semimajor axis in the region $a < 2.15$ AU.

The phenomenon of slow chaotic diffusion of asteroids in the inner main belt is important because it may explain how the Mars-crosser population is kept in a sort of steady state (Migliorini *et al.*, 1998; Morbidelli and Nesvorný, 1999). The population of Mars-crossers, in turn, is believed to be one of the major transient sources of Earth-crossers, in addition to the 3/1 and ν_6 resonances (Bottke *et al.*, 2000b; Michel *et al.*, 2000).

An analogous study of the slow chaotic diffusion of small bodies, but for the Kuiper belt, has recently been done by Nesvorný and Roig (2001). They integrated over 4 Gy the evolution of 101 test particles, with initial semimajor axis evenly distributed in the [37,39] AU range, initial eccentricity and inclination set respectively equal to 0.01 and 2 degrees and randomly chosen phase angles. Figure 12.7 shows the evolution of the proper semimajor axis and eccentricity of these particles, computed as described above for the asteroids. The dark points denote the trajectory before the particle's perihelion distance first becomes smaller than 32 AU; larger gray symbols show the evolution of the trajectory after this instant. In total, 16 test particles reach Neptune-crossing orbit during the simulation. The two leftmost particles are at the borders of the 3/4 mean motion resonance with Neptune (the latter being centered at 36.48 AU), while the rightmost escaping particle is close to the chaotic border of the 2/3 resonance. The remaining 13 particles which become Neptune-crossers evolved from their respective initial locations due to the slow increase of their proper eccentricity, driven by higher-order mean motion resonances with Neptune. Most of them (10) are in the 5/7 resonance and the mean time required to reach Neptune-crossing orbit is 666 My. The others are in the 8/11 (crossing time: 641 My), 7/10 (620 My) and 9/13 (1419 My) mean motion resonances. Some other particles significantly diffuse in proper eccentricity due to three-body resonances with Uranus and Neptune, but do not reach Neptune-crossing orbit within the integration timespan. As for the asteroid case, the proper semimajor axis stays almost constant until the particles start to have close encounters with Neptune, while afterwards it macroscopically changes with time.

A more extended view of the dynamics in the Kuiper belt has been provided by Duncan *et al.* (1995). They integrated over 4 Gy the evolution of a few

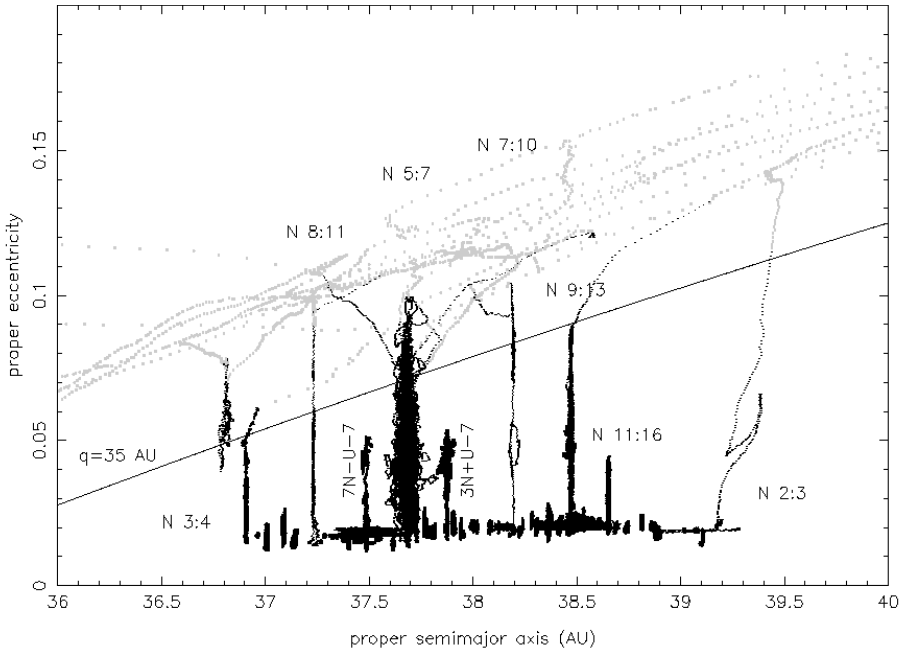


Figure 12.7: The same as the bottom panel of Fig. 12.5, but for test particles in the inner Kuiper belt. The gray dots denote the evolution of the bodies after their perihelion distance has first decreased below 32 AU. The labels denote the main resonances with Neptune and three-body resonances with Uranus and Neptune. Reprinted from Fig. 3 of Nesvorný and Roig (2001), with permission from Academic Press.

thousand fictitious particles, with initial conditions chosen on regular grids on the (a, e) plane at low inclination and on the (a, i) plane at small eccentricity, and a ranging from 32 to 50 AU Figure 12.8 associates to each initial condition its dynamical lifetime. The latter is defined as the time spent before Neptune is first encountered within a distance of 3 Hill's sphere radii. The yellow color denotes the particles that never encounter Neptune within the integration time, but it *does not* imply that these bodies have a regular motion: the motion may also be chaotic, provided that encounters with Neptune do not occur. In this respect, Fig. 12.8 should be considered as a map of the regions that lead to *macroscopic instability*, rather than a map of the chaotic regions.

The main dynamical properties of the Kuiper belt illustrated in Fig. 12.8 are the following. The (a, e) plane is basically divided by the curve $q = 35$ AU into a part which is characterized by lifetimes as long as 4 Gy and a part which is unstable on a much shorter timescale. Above the $q = 35$ AU curve,

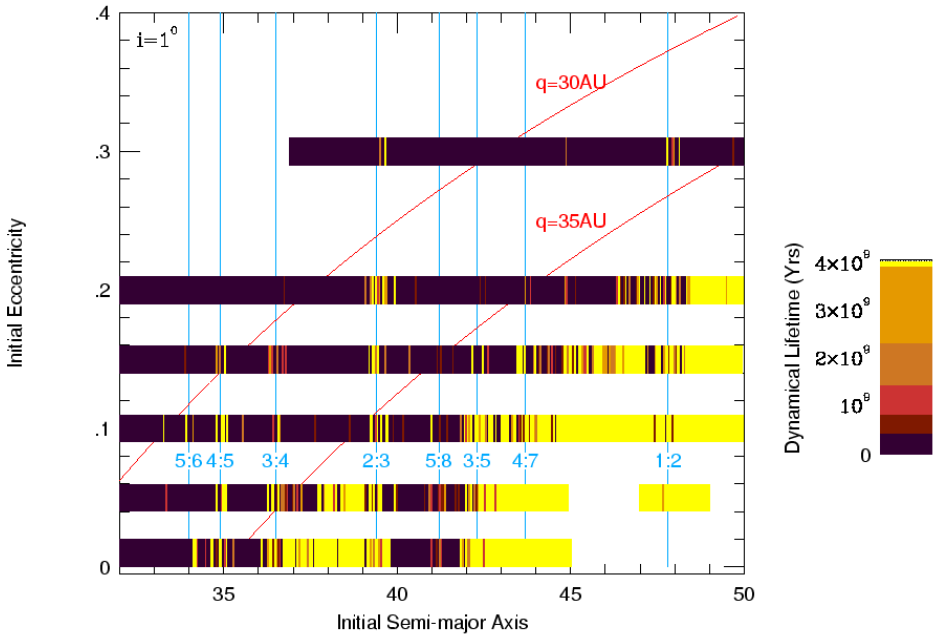


Figure 12.8: The dynamical lifetime of test particles in the Kuiper belt, as a function of their initial semimajor axis and eccentricity. The initial inclination is set to 1 degree. The lifetime is defined as the time required to encounter Neptune within 3 Hill's radii, for the first time. The labels $n : m$ denote the main mean motion resonances with Neptune. Reprinted from Fig. 1 of Duncan *et al.* (1995), with permission from the American Astronomical Society.

the only orbits with long lifetimes are those associated to first-order mean motion resonances, as the latter provide an effective phase protection against close encounters with Neptune (see Section 9.1.1). Below the $q = 35$ AU curve, the main unstable regions (where even particles on initial circular orbits have short dynamical lifetimes) are those between 35–36 AU and 40–42 AU, where the perihelion secular resonances ν_7 and ν_8 are located (see Fig. 8.10). The 40–42 AU instability region divides the Kuiper belt into two parts. In the *inner* belt ($a < 40$ AU), stable orbits are associated with first-order mean motion resonances, or are located in the 36–39 AU region at low eccentricity. In this region – as shown in Fig. 12.7 – only a minority of bodies, associated

with specific tiny resonances, may escape. In the *outer* belt ($a > 42$ AU, also called the *classical* belt) most of the stable orbits are nonresonant. In this respect, the dynamical structure of the Kuiper belt seems to be the mirror image of that of the asteroid belt.

Figure 12.8 also shows the existence of orbits which encounter Neptune after only a few billion years of evolution. Some of these orbits are evidently associated with the main mean motion resonances, and they are presumably related to the slow diffusion regions at moderate amplitude of libration discussed in Section 11.4. The location of the remaining ones correlate quite well with the main peaks of the Lyapunov exponent distribution shown in Fig. 12.3. It is then likely that these orbits slowly diffuse in proper eccentricity, driven by high-order two-body and three-body resonances, analogously to what is shown in Fig. 12.5 for the asteroid belt and in Fig. 12.7 for the inner Kuiper belt.

12.3 Analytic estimates of Lyapunov time and instability time

The previous sections have undoubtedly shown that two-body resonances of quite large eccentricity order and three-body resonances are important for the generation of chaos and for the slow diffusion of the small bodies' eccentricities and inclinations. However, the results illustrated above are purely numeric; it would be important and useful to achieve a theoretical understanding of the phenomenon of weak chaos in mean motion resonances, and to obtain analytic estimates for the Lyapunov exponent and the diffusion timescale.

An important attempt in this direction has been recently made by Murray and Holman (1997). Because it is very difficult (if not impossible) to treat the problem rigorously and in a general way, their work involves a number of approximations; nevertheless it gives estimates that are correct within an order of magnitude in most cases.

As discussed in Chapters 9 and 10, the origin of chaos is attributed to the partial overlapping of the resonances in the mean motion resonant multiplet. To follow Murray and Holman in the notation of this book, we start from Hamiltonian (9.18) and we restrict it to the planar elliptic three-body problem. Namely, of all the harmonics $c_{m,\mathbf{u},\mathbf{v},s,r} \cos \sigma_{m,\mathbf{u},\mathbf{v},s,r}$ in (9.18) we retain only those with $\mathbf{v} = \mathbf{0}, r = 0$ and $u_j = 0$ for $j \neq \bar{\jmath}$, where $\bar{\jmath}$ is the index of the resonant planet. Moreover, for simplicity we consider only the terms with $m = 1$, which sets $u_{\bar{\jmath}} = -(k_{\bar{\jmath}} - k - s)$. With the abridged notation $c_s = c_{1,u_{\bar{\jmath}},\mathbf{0},s,0}$

and $c_{\mathbf{0}} = c_{0,\mathbf{0},\mathbf{0},0,0}$, our model Hamiltonian becomes:

$$\mathcal{H}_{\mathrm{R3B}} = \mathcal{H}_0(\Lambda, \Lambda_{\bar{\jmath}}) + c_{\mathbf{0}}(\Lambda, P) + \sum_{s=0}^{k_{\bar{\jmath}}-k} c_s(\Lambda, P) \cos(k_{\bar{\jmath}}\lambda_{\bar{\jmath}} - k\lambda + sp - (k_{\bar{\jmath}} - k - s)\varpi_{\bar{\jmath}}) \ , \tag{12.2}$$

where $\varpi_{\bar{\jmath}}$ is the perihelion longitude of the resonant planet. The latter is assumed to be fixed, so that the part of $\mathcal{H}_0$ that depends on $\Lambda_{g_j}, \Lambda_{s_j}$ (see formula 9.16) has been dropped. For simplicity, we will assume below that $\varpi_{\bar{\jmath}} = 0$, so we drop it from (12.2). The coefficients c_s are all of order $|k_{\bar{\jmath}} - k|$ in the sum of the powers of the eccentricities of the small body and of the resonant planet. We now define the canonical variables

$$\begin{aligned} \Psi &= (\Lambda - \Lambda_0)/k \ , & \psi &= k\lambda - k_{\bar{\jmath}}\lambda_{\bar{\jmath}} \\ I &= P \ , & \varphi &= p \\ \tilde{\Lambda}_{\bar{\jmath}} &= \Lambda_{\bar{\jmath}} + k_{\bar{\jmath}}(\Lambda - \Lambda_0)/k \ , & \tilde{\lambda}_{\bar{\jmath}} &= \lambda_{\bar{\jmath}} \ , \end{aligned} \tag{12.3}$$

where Λ_0 is the unperturbed location of the mean motion resonance given by Kepler's law. Retaining from the Taylor expansion of $\mathcal{H}_0$ and c_0 in (12.2) only the lowest-order terms in Ψ, I, we get:

$$\mathcal{H}_{\mathrm{R3B}} = \frac{1}{2}\beta\Psi^2 + 2\varepsilon A I + \sum_{s=0}^{k_{\bar{\jmath}}-k} c_s(I)\cos(\psi - s\varphi) \ , \tag{12.4}$$

where $\beta = \partial^2\mathcal{H}_0/\partial^2\Lambda(\Lambda_0, \Lambda_{\bar{\jmath}})$, $2\varepsilon A = \partial c_0/\partial I(\Lambda_0, 0)$ and $c_s(I) = c_s(\Lambda_0, I)$. Although derived in the simple framework of the restricted three-body problem for a two-body resonance, (12.4) can be considered as a more general planar model of the multiplet structure of a mean motion resonance, that is valid also for three-body resonances. In this respect, the coefficient $2\varepsilon A$ should be regarded as the frequency of precession of the secular angles, which is explicitly indicated to be of order ε (the mass of the planet that dominates the secular dynamics); the coefficients c_s are instead proportional to the mass of the resonant planet for two-body resonances, or the product of the masses of the resonant planets for three-body resonances, and they are typically smaller than ε; the sum of the harmonic terms should be extended to all combinations of the secular angles that give resonant harmonics of the lowest possible eccentricity order.

To obtain a general model that can be handled analytically, Murray and Holman assume that all coefficients c_s are equal to the leading one c_{s_0}, and

that the sum over s is extended from $-\infty$ to $+\infty$, thus obtaining:

$$\begin{aligned}\mathcal{H}_{\rm R3B} &= \frac{1}{2}\beta\Psi^2 + 2\varepsilon AI + c_{s_0}(I)\cos(\psi)\sum_{s=-\infty}^{+\infty}\cos(s\varphi)\\ &= \frac{1}{2}\beta\Psi^2 + 2\varepsilon AI + c_{s_0}(I)\cos(\psi)\delta(\varphi) \qquad (12.5)\end{aligned}$$

where $\delta(\varphi)$ is Dirac's delta function with periodic argument φ. Integrating the equations given by (12.5) for Ψ and ψ over a period of φ one gets the symplectic mapping (see Chirikov, 1979; Wisdom, 1983)

$$\Psi' = \Psi + \frac{\pi c_{s_0}(I)}{\varepsilon A}\sin\psi \ , \quad \psi' = \psi + \frac{\pi\beta}{\varepsilon A}\Psi' \ . \qquad (12.6)$$

Recalling now from (12.4) that $\dot I \sim -s_0\dot\Psi$ one gets the supplementary map for the evolution of the action I:

$$I' = I - s_0\frac{\pi c_{s_0}(I)}{\varepsilon A}\sin\psi \ . \qquad (12.7)$$

With the rescaling $\bar\Psi = \Psi\pi\beta/(\varepsilon A)$, the mapping (12.6) is analogous to the standard map in the form

$$\bar\Psi' = \bar\Psi + K_{\rm eff}\sin\psi \ , \quad \psi' = \psi + \bar\Psi' \ , \qquad (12.8)$$

where

$$K_{\rm eff}(I) \equiv \left(\frac{c_{s_0}(I)}{\beta}\right)\left(\frac{\pi\beta}{\varepsilon A}\right)^2 \ . \qquad (12.9)$$

It is well known (see Chapter 4 in Lichtenberg and Lieberman, 1983 and Section 6.1) that the standard map (12.8) exhibits large-scale chaos if $K_{\rm eff} > K_{\rm crit}$, with $K_{\rm crit} \sim 1$. Notice that the condition $K_{\rm eff} > 1$ is similar to the condition for the overlapping of the resonances in the resonant multiplet of model (12.4); in fact, the separation among the resonances is $\delta\Psi = 2\varepsilon A/\beta$ while their width is $\Delta\Psi \sim 2\sqrt{c_{s_0}/\beta}$, so that $K_{\rm eff} = (\pi\Delta\Psi/\delta\Psi)^2$. The factor π^2 between $K_{\rm eff}$ and $(\Delta\Psi/\delta\Psi)^2$ accounts for the fact that, as explained in Section 6.1, the transition to global chaos occurs *before* the low-order resonances in the multiplet overlap, due to the existence of overlapping higher-order resonances in between them.

To estimate the value of the Lyapunov exponent, Murray and Holman write the tangent map for the mapping (12.6):

$$\delta\Psi' = \delta\Psi + \delta\psi\frac{\pi c_{s_0}(I)}{\varepsilon A}\cos\psi \ , \quad \delta\psi' = \delta\psi + \frac{\pi\beta}{\varepsilon A}\delta\Psi' \ , \qquad (12.10)$$

whose eigenvalues are given by the equation

$$(1-\lambda)(1+K_{\rm eff}\cos\psi-\lambda)-K_{\rm eff}\cos\psi=0\ , \qquad (12.11)$$

with $K_{\rm eff}$ given by (12.9). Solving (12.11) for λ one gets

$$\lambda=1+\frac{1}{2}K_{\rm eff}\cos\psi\pm\sqrt{K_{\rm eff}\cos\psi+\left(\frac{1}{2}K_{\rm eff}\cos\psi\right)^2}\ . \qquad (12.12)$$

In a chaotic layer, it is reasonable to expect that ψ ergodically assumes all possible values between 0 and 2π, so that the Lyapunov exponent of the mapping can be estimated by averaging the logarithm of λ over ψ:

$$\gamma=\frac{1}{2\pi}\int_0^{2\pi}\ln\lambda{\rm d}\psi\ . \qquad (12.13)$$

Murray and Holman approximate[1]

$$\gamma\sim\ln\left[1+\frac{K_{\rm eff}}{4}+\sqrt{\frac{K_{\rm eff}}{2}+\left(\frac{K_{\rm eff}}{4}\right)^2}\right]\ . \qquad (12.14)$$

As the period of the map (12.6) is $\pi/(\varepsilon A)$ (the period of φ in the Hamiltonian 12.4), the Lyapunov time of the considered mean motion resonance will be

$$T_{\rm L}=\frac{\pi}{\varepsilon A}\frac{1}{\gamma}\sim\frac{\pi}{\varepsilon A}\Bigg/\ln\left[1+\frac{K_{\rm eff}}{4}+\sqrt{\frac{K_{\rm eff}}{2}+\left(\frac{K_{\rm eff}}{4}\right)^2}\right]\ . \qquad (12.15)$$

To test the validity of this estimate, Murray and Holman have performed a series of numerical tests. Figure 12.9 shows the Lyapunov time of test bodies in the elliptic planar restricted three-body problem in various mean motion resonances with Jupiter in the outer asteroid belt (the initial eccentricity was chosen equal to 0.05). The solid squares give the Lyapunov times as measured in the numerical integrations, while the predictions (12.15) are shown as open circles. As one sees, the agreement is generally good within an order of magnitude, although a few exceptions can be noted.

To estimate the diffusion of the eccentricity of bodies in mean motion resonances, Murray and Holman start from the known result (see Chapter 5 of Lichtenberg and Lieberman, 1983) that, when $K_{\rm eff}>K_{\rm crit}$ in (12.8), the transport of the action $\bar{\Psi}$ can be described by the Fokker–Plank equation. This means that, given a large ensemble of particles, their distribution with

[1] There is a missprint in the corresponding formula given in Murray and Holman (1997).

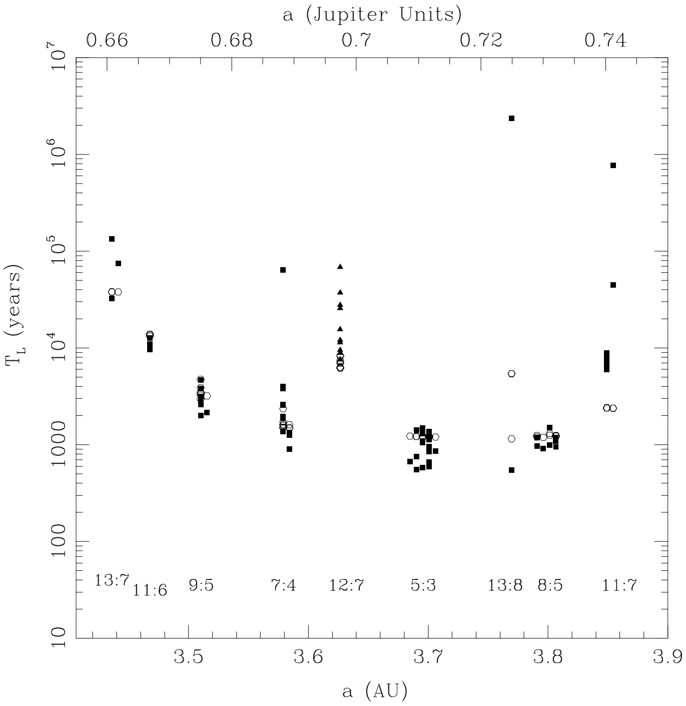


Figure 12.9: Lyapunov time as a function of the semimajor axis in the outer asteroid belt, in the framework of the elliptic restricted three-body problem. The labels $k_j : k$ denote the major mean motion resonances with Jupiter. Solid symbols are numerical measurements: triangles report data obtained through integrations including the effects of the four giant planets, while squares are obtained taking into account the sole effect of Jupiter. The circles represent the predictions of formula (12.15). Reprinted from Fig. 1 of Murray and Holman (1997), with permission from the American Astronomical Society.

respect to $\bar{\Psi}$ can be represented by a function $P(\bar{\Psi}, t)$ evolving under the equation

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial \bar{\Psi}} \left(\frac{D}{2} \frac{\partial P}{\partial \bar{\Psi}} \right) , \tag{12.16}$$

where $D = K_{\text{eff}}^2/2$ is called the *diffusion coefficient*. Recalling the scaling relationship $\bar{\Psi} = \Psi\pi\beta/(\varepsilon A)$, the diffusion coefficient for the action Ψ becomes

$$D_\Psi = \frac{K_{\text{eff}}^2}{2} \left(\frac{\varepsilon A}{\pi\beta} \right)^2 = \frac{1}{2} \left(\frac{c_{s_0}}{\beta} \right)^2 \left(\frac{\pi\beta}{\varepsilon A} \right)^2 . \tag{12.17}$$

Now taking into account that the timescale for the mapping (12.6) is $\pi/(\varepsilon A)$,

one finally obtains that the diffusion coefficient for Ψ is

$$D_\Psi = \frac{1}{2} c_{s_0}^2(I) \frac{\pi}{\varepsilon A} \ . \tag{12.18}$$

Because the map (12.7) for the evolution of I differs from that for Ψ by a factor s_0, the action I will respond to a Fokker–Plank equation with diffusion coefficient

$$D_I = \frac{s_0^2}{2} c_{s_0}^2(I) \frac{\pi}{\varepsilon A} \ . \tag{12.19}$$

Therefore, in Murray and Holman's model the resonant particles random walk in both the semimajor axis (or Ψ) and eccentricity (or I). However, both random walks are subject to certain boundary conditions. The semimajor axis is restricted to lie in the interval defined by the overlap of the components of the resonant multiplet, which in reality does not have an infinite extension. This can be modeled by reflecting boundaries at specific values of Ψ. The eccentricity is restricted to be positive, which can be modeled by a reflecting boundary at $I = 0$. Because of this reflecting boundary, test particles on chaotic orbits at small initial I diffuse from small eccentricities to large eccentricities. When they reach an eccentricity that is large enough to cross the orbit of a planet, they are removed by close encounters. This is modeled by an absorbing boundary at $I = I_{\rm max}$. Solving the Fokker–Plank equation, Murray and Holman computed that the typical *removal time*, that is the timescale for particles started at I_0 to reach $I_{\rm max}$, is

$$T_{\rm R} \sim \frac{I_{\rm max} I_0}{\sqrt{D_I(I_{\rm max}) D_I(I_0)}} \ . \tag{12.20}$$

Again, Murray and Holman tested the validity of their estimate with numerical simulations. Fig. 12.10 is the analog of Figure 12.15 but compares the predicted with the observed removal times (respectively indicated by circles and filled squares). Because in (12.20) $T_{\rm R}$ depends strongly on the eccentricity used to evaluate I_0, Murray and Holman used the average eccentricity computed from the numerical integrations over the first few thousand orbital periods. The scatter in the values of I_0 thus obtained accounts for the scatter of the predicted removal times in each resonance, shown in Fig. 12.10. Some caution must also be paid to evaluate $D_I(I_{\rm max})$. In fact if one chooses $I_{\rm max}$ from the value of the eccentricity that allows a body to cross the orbit of Jupiter, the coefficient $c_{s_0}(I_{\rm max})$ is infinite, because the Hamiltonian is not analytic and cannot be expanded in a Fourier series as in (12.2). As a consequence, also $D_I(I_{\rm max})$ is infinite and $T_{\rm R}$ is zero. Therefore, one has to choose a somewhat smaller value of $I_{\rm max}$. To remove the arbitrariness of this choice, the best strategy is to choose $I_{\rm max}$ so as to maximize the value of $T_{\rm R}$.

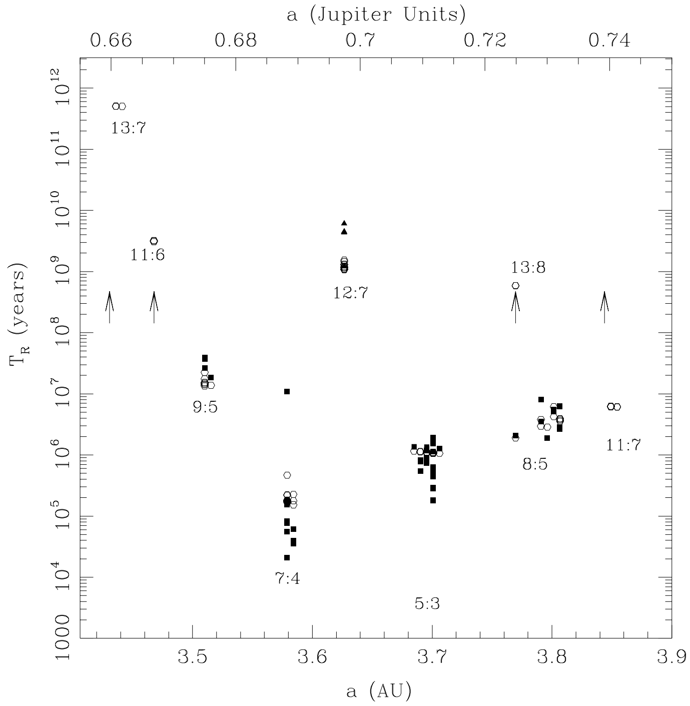


Figure 12.10: The same as Fig. 12.9, but for the removal time. Solid symbols are numerical measurements, while circles represent the predictions of formula (12.20). Arrows indicate the integration timespan for those resonances where none of the particles has been removed. Reprinted from Fig. 2 of Murray and Holman (1997), with permission from the American Astronomical Society.

As one sees from Fig. 12.10, the agreement between predictions and numerical tests is generally good within an order of magnitude, although some exceptions can be noted. This is remarkable, given the large number of approximations involved to achieve the final estimates. We stress that the diffusion coefficient (12.19) for the action I does not take into account the possible effects of the purely secular dynamics, which – as we have seen in Chapter 11 – may force the eccentricity to increase in a shorter timescale. Moreover, the assumption made at the beginning of this section, that approximates the resonant multiplet by an infinite chain of resonances, is not very suitable for resonances of low eccentricity order, which behave as a modulated pendulum (see Section 9.4). In this case, still neglecting the important effects of the purely secular dynamics, the diffusion coefficients for orbits in the chaotic layer would be simply proportional to the cube of the modulation frequency,

i.e. $(\varepsilon A/\pi)^3$ (Bruhwiler and Cary, 1989), i.e. smaller than (12.18) and (12.19). This might explain some of the major discrepancies visible in Fig. 12.10.

Despite the limitations, Murray and Holman's estimate of the removal time is an important attempt to go beyond classical perturbation theory and cope with the statistical properties of chaotic evolution.

12.4 Do the KAM and Nekhoroshev theorems apply for small body dynamics?

At this point, it is natural to ask whether the view of the global dynamics provided by the KAM and the Nekhoroshev theorems (see Chapters 3 and 6 respectively) is suitable to describe small body dynamics in the framework of our Solar System.

Of course both theorems cannot be straightforwardly applied, as they are designed for *nondegenerate* dynamical systems, but they must be first adapted to the case of the perturbed Kepler motion. Arnold (1963c) and Nekhoroshev (1977) already worked in this direction, and, more recently, Guzzo and Morbidelli (1997) reformulated both theorems taking into account most features of small body dynamics.

It should be evident from the previous chapters of this book that, whatever the values of the planetary masses and eccentricities, there is no hope of proving a Nekhoroshev-like stability result that holds over all the non-planet-crossing phase space. In fact, the mean motion resonances whose eccentricity order and Fourier order[2] are both low to moderate (the threshold order depending on the small body's eccentricity) generically give rise to a multiplet of resonances which overlap (see Section 9.3). In this case, it is possible to show with a plausible model (see Section 12.3) that macroscopic changes of the small body's eccentricity may occur in a *short time*, i.e. in a time that is proportional to some power of the perturbation parameters (see formula 12.20). It is true that the resonances of lowest order behave as a modulated pendulum (see Section 9.4), and therefore have a central island where the motion might be stable, but we have seen in Chapter 11 that in these cases the stability or instability properties ultimately depend on the specific features of the secular dynamics, which vary from resonance to resonance. Therefore, it might be possible to prove a Nekhoroshev-like stability result restricted to some specific resonance (for instance restricted to the central core of the 3/2 resonance with Jupiter), but it is not possible to achieve a general result valid for all resonances of low order.

[2]See Section 9.1 for their definition.

As a consequence, a general Nekhoroshev-like stability result might be established only on the nonconnected domain that excludes all the mean motion resonances with low to moderate order. This domain exists only if the low to moderate-order resonances do not overlap each other, which gives a condition on the maximal value ε of the planetary masses, on the minimal distance from the resonant planet and on the maximal value of the small body's eccentricity. For reference we will call this domain the *Nekhoroshev domain*.

A detailed description of how a Nekhoroshev result can be established goes beyond the scope of this section, and we report just a brief guideline. The interested reader can refer to Guzzo and Morbidelli (1997). In the Nekhoroshev domain, one first considers the subdomains where there are no mean motion resonances of Fourier order smaller than $1/\varepsilon$. On these subdomains one can construct the secular normal form with a remainder that is exponentially small in $1/\varepsilon$, and restrict the study of the stability to the secular problem. The latter being nondegenerate, the Nekhoroshev theorem can be applied in its classical form (see Chapter 6). A technical complication is that the condition (6.2) does not hold for the secular Hamiltonian, but in general one can bound the secular resonant motion using the third-order derivatives in the actions. In this way, one can still obtain a stability time that is exponentially long in $1/\eta$, where η is some small parameter – typically related to the planetary eccentricities and inclinations. The ν_6 secular resonance is an exception, and indeed the eccentricity of ν_6 resonant bodies can grow from 0 to 1 on a "short" timescale, as shown in Section 8.4.2.

Elsewhere in the Nekhoroshev domain, one has to cope with mean motion resonances of nonnegligible order (i.e. smaller than $1/\varepsilon$). However, by definition of the Nekhoroshev domain, these resonances have a resonant multiplet in which the resonant components do not overlap. Therefore, with a procedure analogous to that sketched in Section 6.2.1 one can still achieve a stability result, also valid for a time that is exponentially long in $1/\eta$.

Once a Nekhoroshev-like stability result is proved over the Nekhoroshev domain, the procedure can be iterated as discussed in Section 6.2.2, thus proving also the existence of KAM tori. Therefore, the Nekhoroshev-like stability result is intimately related to the existence of a specific dynamical structure, analogously to the case of nondegenerate systems.

These results show that a suitable adaptation of the KAM and Nekhoroshev theorems can be applied, in principle, to describe the small body dynamics, at least in parts of the phase space. However, it is beyond current possibilities to analytically determine if these theorems do really apply in the framework of our Solar System (characterized by specific values of the planetary masses, semimajor axes, eccentricities and inclinations), and in which

regions. However, as explained in Section 6.2.2, the numerical simulations are very useful to understand whether the dynamical structure of a given system is that described by these theorems. Although a detailed analysis has never been done (it would require the study of how the dynamical structure changes with increasing accuracy of detection of the chaotic zones), Fig. 12.1 suggests that if only the four giant planets are taken into account, the inner and central part of the asteroid belt might have the dynamical structure described by the Nekhoroshev theorem, at least at moderate eccentricity. The outer belt, conversely, is much more perturbed by Jupiter, so that the chaotic regions related to two-body and three-body mean motion resonances rapidly overlap with increasing asteroid eccentricity (see Fig. 10.8). When the inner planets are taken into account, Fig. 12.2 does not leave much hope that the Nekhoroshev structure might persist in the inner belt (at least up to 2.15 AU), the mean motion resonances with Mars densely filling the region. Therefore, in the framework of the complete planetary system one should expect that the Nekhoroshev structure is present only in the central belt, at small eccentricity. This is also an indication of the frequency map analysis illustrated in Fig. 12.4. The same figure also shows that nowhere else in the inner Solar System and in the giant planet region should one expect the Nekhoroshev theorem to hold for small body dynamics. Only in the Kuiper belt, far enough from Neptune, can one find again a Nekhoroshev-like structure. Between 43 and 50 AU the Nehoroshev structure is confined at very small eccentricity, while beyond 50 AU a more considerable region seems to be dominated by regular motion, segmented by well-separated individual resonances. It is possible that a more accurate numerical investigation would show that, of the entire Solar System, only the deep Kuiper belt is characterized by a real Nekhoroshev structure.

Notice, however, that the search for a Nekhoroshev structure is more a mathematical curiosity than an astronomical exigency. The stability times associated with the Nekhoroshev structure are likely to exceed the physical lifetime of the solar system by many orders of magnitude. Small bodies can therefore live for the age of the Solar System also in mean motion resonances that are characterized by an overlapping resonant multiplet, providing that the resulting removal time (12.20) is long enough. For instance, this seems to be the case of the asteroid Helga in the 12/7 resonance with Jupiter, whose removal time, from analytic estimates and numerical integrations, is about 5 Gy (Murray and Holman, 1997).

The problem of the possible validity of a Nekhoroshev-like stability result is also complicated by the fact that the perturbing planets themselves exhibit chaotic behavior (see Chapter 7 and Section 10.4). This must destroy at some level the Nekhoroshev structure of small body dynamics, and prevent the exis-

tence of invariant tori, as can be understood as follows. A non-quasi-periodic perturbation has a continuous Fourier spectrum. In the case of weak chaos, such as that of the planets, one can reasonably assume that the continuous Fourier spectrum is nonzero only on bounded intervals of width $\Delta\omega$ around the discrete values of the frequencies characterizing the quasi-periodic approximation of their motion. As a consequence of the continuous spectrum, each resonance splits into a continuous package of resonances. The width of this package in the space of frequencies is $\Delta\omega$, whatever the order of the resonance is. Then, the width of each resonance domain must be larger than $\Delta\omega$, instead of decaying exponentially with the order of the resonance as in the usual case. Referring to formula (6.5) the volume filled by resonances of order K becomes larger than

$$V_K \sim 2^n K^{n-1} \left[\sqrt{\varepsilon} \exp\left(-K\sigma\right) + \Delta\omega\right] \ , \qquad (12.21)$$

which indefinitely increases with K. Therefore, at some order K resonances must overlap, thus destroying the Nekhoroshev structure and the existence of invariant tori. Consequently, if the numerical analysis on asteroid dynamics is pushed to high enough accuracy, eventually a chaotic continuum should be detected even in the regions that seem to be the most regular ones. Whatever the initial conditions, the maximum Lyapunov exponent should be positive.

However, in the case of our Solar System, the size of the intervals $\Delta\omega$, defined with respect to some cut-off level, is also quite small, since the frequencies of the inner planets (which are the most chaotic ones) have a relative change of a few percent only (see Section 7.3). Therefore, the effects of chaotic perturbations might not be relevant to the stability of small body motion over the age of the Solar System.

References

Agnor C.B., Canup R.M. and Levison H.F. (1999) On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus*, **142**, 219.

Applegate J.H., Douglas M.R., Gursel Y., Sussman G.J. and Wisdom J. (1986) The outer solar system for 200 million years. *Astron. J.*, **92**, 176.

Arnold V.I. (1963a) On a theorem of Liouville concerning integrable problems of dynamics. *Sib. Mathem. Zh.*, **4**, 2.

Arnold V.I. (1963b) Proof of A.N. Kolmogorov's theorem on the conservation of conditionally periodic motions with a small variation in the Hamiltonian. *Russian Math. Surv.*, **18**, 9.

Arnold V.I. (1963c) Small denominators and problems of stability of motion in classical and celestial mechanics. *Russ. Math. Surveys*, **18**, 85.

Arnold V.I. and Avez A. (1968) *Ergodic Problems of Classical Mechanics.* Benjamin, New York.

Arnold V.I., Kozlov V.V. and Neishtadt A.I. (1988) Mathematical aspects of classical and celestial mechanics, *Dyn. Systems III*, V.I. Arnold, ed. Springer, Berlin.

Bailey M.E., Chambers J.E. and Hahn G. (1992) Origin of Sun-grazers: a frequent cometary end-state. *Astron. Astrophys.*, **257**, 315.

Beaugé C. (1994) Asymmetric liberations in exterior resonances. *Celest. Mech.*, **60**, 225.

Bendjoya Ph., Slezak E. and Froeschlé C. (1991) The wavelet transform: a new tool for asteroid family determination. *Astron. Astrophys.*, **251**, 312.

Benettin G., Galgani L. and Strelcyn J.M. (1976) Kolmogorov entropy and numerical experiments. *Phys. Rev. A*, **14**, 2338.

Benettin G., Galgani L., Giorgilli A. and Strelcyn J.M. (1980) All Lyapunov characteristic exponents are effectively computable. *Meccanica*, March, 21.

Benettin G., Galgani L. and Giorgilli A. (1985) A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems. *Celest. Mech.*, **37**, 1.

Born M. (1927) *The Mechanics of the Atom.* Bell, London.

Bottke W.F., Rubincam D.P. and Burns J.A. (2000a) Dynamical evolution of main belt meteoroids: numerical simulations incorporating planetary perturbations and Yarkovsky thermal forces. *Icarus*, **145**, 301.

Bottke W.F., Jedicke R., Morbidelli A., Petit J.M. and Gladman B. (2000b) Understanding the distribution of Near Earth Asteroids. *Science*, **288**, 2190.

Bretagnon P. (1974) Termes à longues périodes dans le système solaire. *Astron. Astrophys.*, **30**, 141.

Bretagnon P. (1982) Theory for the motion of all the planets: the VSOP82 solution. *Astron. Astrophys.*, **114**, 278.

Bretagnon P. (1984) Improvement of analytic planetary theories. *Celest. Mech.*, **34**, 193.

Bretagnon P. (1990) An iterative method for the construction of a general planetary theory. *Astron. Astrophys.*, **231**, 561.

Brouwer D. and Clemence G.M. (1961) *Methods of Celestial Mechanics.* Academic Press, New York.

Bruhwiler D.L. and Cary J.R. (1989) Diffusion of particles in a slowly modulated wave. *Physica D*, **40**, 265.

Cameron A.G. (1997) The origin of the Moon and the single impact hypothesis. *Icarus*, **126**, 126.

Canup R.M. and Esposito L.W. (1996) Accretion of the Moon from an impact-generated disk. *Icarus*, **119**, 427.

Carpino M., Milani A. and Nobili A.M. (1987) Long-term numerical integrations and synthetic theories for the motion of the outer planets. *Astron. Astrophys.*, **181**, 182.

Cellino A., Michel P., Tanga P., Zappalà V., Paolicchi P. and dell'Oro A. (1999) The velocity–size relationship for members of asteroid families and implications for the physics of catastrophic collisions. *Icarus*, **141**, 79.

Chambers J.E. (1999) A hybrid symplectic integrator that permits close encounters between massive bodies. *Month. Not. Roy. Astron. Soc.*, **304**, 793.

Chambers J.E. and Wetherill G.W. (1998) Making the terrestrial planets: N-body integrations of planetary embryos in three dimensions. *Icarus*, **136**, 304.

Chambers J.E., Wetherill G.W. and Boss A.P. (1996) The stability of multiplanet systems. *Icarus*, **119**, 261.

Chirikov B.V. (1960) Resonance processes in magnetic traps. *Plasma Phys.*, **1**, 253.

Chirikov B.V. (1979) A universal instability of many dimensional oscillator systems. *Physics Reports*, **52**, 265.

Cincotta P. and Giordano C.M. (2001) Global dynamics through the Mean Exponential Growth factor of Nearby Orbits (MEGNO). Preprint of Obs. of La Plata, Argentina.

Cincotta P. and Simó C. (2000) Simple tools to study global dynamics in non-axisymmetric galactic potentials. *Astron. Astrophys. Suppl.*, **147**, 205.

Colombo G. (1966) Cassini's second and third laws. *Astron. J.*, **71**, 891.

Contopoulos G. (1966) Resonance phenomena and the non-applicability of the "third" integral. in *Les Nouvelles Méthodes de la Dynamique Stellaire*, Nahon F. and Hénon M., eds. CNRS, Paris. See also *Bull. Astron.*, Ser. 3, **2**, 233 (1967).

Contopoulos G. and Voglis N. (1996) Spectra of stretching numbers and helicity angles in dynamical systems. *Celest. Mech.*, **64**, 1.

Contopoulos G., Voglis N., Efthyomopoulos C., Froeschlé C., Gonczi R., Lega E., Dvorak R. and Lohinger E. (1997) Transition spectra of dynamical systems. *Celest. Mech.*, **67**, 293.

Danby J.M.A. (1962) *Fundamentals of Celestial Mechanics.* Willmann-Bell Inc. Richmond, Virginia.

De la Barre C.M., Kaula W.M. and Varadi F. (1996) A study of orbits near Saturn's triangular Lagrangian points. *Icarus*, **121**, 88.

Delaunay C. (1867) Théorie du mouvement de la Lune. *Mem. Acad. Sci.*. **29**, Paris.

Delshams A., Ramirez-Ros R. and Seara T.M. (1999) Splitting of separatrices in Hamiltonian systems and symplectic maps. In *Hamiltonian Systems with Three or More Degrees of Freedom*, C. Simó, ed. Kluwer, Dordrecht, 39.

Deprit A. (1968) Canonical transformations depending on a small parameter. *Cest. Mech.*, **1**, 12.

Dermott F.S. and Murray C.D. (1983) Nature of the Kirkwood gaps in the asteroid belt. *Nature*, **301**, 201.

Duncan M.J., Quinn T. and Tremaine S. (1987) The formation and extent of the solar system comet cloud. *Astron. J.*, **94**, 1330.

Duncan M.J., Levison H.F. and Budd S.M. (1995) The long-term stability of orbits in the Kuiper belt. *Astron. J.*, **110**, 3073.

Duncan M.J., Levison H.F. and Lee M.H. (1998) A multiple time step symplectic algorithm for integrating close encounters. *Astron. J.*, **116**, 2067.

Duriez L. (1979) 'Approche d'une théorie générale planétaire en variables elliptiques héliocentriques'. Ph.D. thesis, Univ. de Lille, France.

Duriez L. (1989) Le développement de la fonction perturbatrice. in *Les Méthodes Modernes de la Mécanique Céleste*, Benest D. and Froeschlé C., eds. Ed. Frontières, Gif sur Yvettes, France.

Ellis K.M. and Murray C.D. (2000) The disturbing function in solar system dynamics. *Icarus*, **147**, 129.

Elskens Y. and Escande D.F. (1991) Slowly pulsating separatrices sweep homoclinic tangles where islands must be small: an extension of classical adiabatic theory. *Nonlinearity*, **4**, 615.

El-Zant A.A. (1997) On the stability of motion of N-body systems: a geometric approach. *Astron. Astrophys.*, **326**, 113.

El-Zant A.A. and Gurzadyan V.G. (1998) Relative chaos in gravitating systems with massive centre. *Physica D*, **122**, 241.

Epstein P.S. (1916) Zur Quantentheorie, *Ann. d. Physik*, **51**, 168.

Farinella, P. and Vokrouhlický D. (1998) Semimajor axis mobility of asteroidal fragments, *Science*, **283**, 1507.

Farinella P., Froeschlé Ch. and Gonczi R. (1993a) Meteorites from the asteroid 6 Hebe. *Celest. Mech.*, **56**, 287.

Farinella P., Gonczi R., Froeschlé Ch. and Froeschlé C. (1993b). The injection of asteroid fragments into resonance. *Icarus*, **101**, 174.

Farinella P., Froeschlé Ch., Froeschlé C., Gonczi R., Hahn G., Morbidelli A. and Valsecchi G.B. (1994) Asteroids falling onto the Sun. *Nature*, **371**, 315.

Fernández J.A. and Ip W.H. (1984) Some dynamical aspects of the accretion of Uranus and Neptune: the exchange of orbital angular momentum with planetesimals. *Icarus*, **58**, 109.

Ferraz-Mello S. (1987) Averaging the elliptic asteroidal problem near a first-order resonance. *Astron. J.*, **94**, 208.

Ferraz-Mello S. (1994) Dynamics in the asteroid 2/1 resonance. *Astron. J.*, **108**, 2330.

Ferraz-Mello S. (1996) A symplectic mapping approach to the study of the stochasticity in asteroidal resonances. *Celest Mech.*, **65**, 421.

Ferraz-Mello S. and Klafke J.C. (1991) A model for the study of the very high eccentricity asteroidal motion: the 3:1 resonance. In *Predictability, Stability and Chaos in N-body dynamical systems*, A.E. Roy, ed. Plenum Press, New York, 177.

Ferraz-Mello S. and Sato M. (1989) The very-high-eccentricity asymmetric expansion of the disturbing function near resonances of any order. *Astron. Astrophys.*, **225**, 541.

Ferraz-Mello S., Nesvorný D. and Michtchenko T.A. (1997) On the lack of asteroids in the Hecuba gap. *Celest. Mech.*, **69**, 171.

Ferraz-Mello S., Michtchenko T.A. and Roig F. (1998a) The determinant role of Jupiter's great inequality in the depletion of the Hecuba gap. *Astron. J.*, **116**, 1491.

Ferraz-Mello S., Nesvorný D. and Michtchenko T.A. (1998b) Chaos, diffusion, escape and permanence of resonant asteroids in gaps and groups. In *Solar System Formation and Evolution*, D. Lazzaro, R. Vieira Martins, S. Ferraz-Mello, J. Fernandez and C. Beaugé, eds. Astron. Soc. of the Pacific, Conf. Series **149**, 65.

Forzoni Accolti E. (1995) Ph.D. thesis, University of Turin, Italy.

Franklin F.A. (1994) An examination of the relationship between chaotic orbits and the Kirkwood gap at the 2:1 resonance. *Astron. J.*, **107**, 1890.

Froeschlé Ch. and Scholl H. (1986) The secular resonance ν_6 in the asteroidal belt. *Astron. Astrophys.*, **166**, 326.

Froeschlé C. and Lega E. (1998) Twist angles: a method for distinguishing islands, tori and weak chaotic orbits. Comparison with other methods of analysis. *Astron. Astrophys.*, **334**, 355.

Froeschlé Ch., Morbidelli A. and Scholl H. (1991) Complex dynamical behaviour of the asteroid 2335 James associated with the secular resonances ν_5 and ν_{16}: numerical studies and theoretical interpretation. *Astron. Astrophys.*, **24**, 553.

Froeschlé C., Gonczi R. and Lega E. (1997) The fast Lyapunov indicator: a simple tool to detect weak chaos. Application to the structure of the main asteroidal belt. *Planet. Space Sci.*, **45**, 881.

Froeschlé C., Guzzo M. and Lega E. (2000) Graphical evolution of the Arnold web: from order to chaos. *Science*, **289**, 2108.

Gantmacher F. (1975) *Lectures in Analytical Mechanics.* MIR, Moscow.

Giorgilli A. and Galgani L. (1985) Rigorous estimates for the series expansions of Hamiltonian perturbation theory. *Celest. Mech.*, **37**, 95.

Giorgilli A. and Locatelli U. (1997) Kolmogorov theorem and classical perturbation theory. *Z. App. Math. Phys.*, **48**, 220.

Giorgilli A. and Morbidelli A. (1997) Invariant KAM tori and global stability for Hamiltonian systems. *Z. App. Math. Phys.*, **48**, 102.

Gladman B., Quinn D., Nicholson Ph. and Rand R. (1996) Synchronous locking of tidally evolving satellites. *Icarus*, **122**, 166.

Gladman B., Migliorini F., Morbidelli A., Zappalà V., Michel P., Cellino A., Froeschlé Ch., Levison H.F., Bailey M. and Duncan M.J. (1997) Dynamical lifetimes of objects injected into asteroid main belt resonances, *Science*, **277**, 197.

Green J.M. (1979) A method for determining stochastic transitions. *J. Math. Phys.*, **20**, 1183.

Gröbner W. (1960) *Die Lie–Reihen und ihre anwendungen.* Springer-Verlag, Berlin.

Gronchi G.F. and Milani A. (1998) Averaging on Earth-crossing orbits *Celest. Mech.*, **71**, 109.

Gronchi G.F. and Milani A. (1999) The stable Kozai state for asteroids and comets with arbitrary semimajor axis and inclination. *Astron. Astrophys.*, **341**, 928.

Gurzadyan V.G. and Kocharyan A.A. (1987) Relative chaos in stellar systems. *Astr. Sp. Sci.*, **135**, 307.

Gurzadyan V.G. and Ruffini R. (2000) *The Chaotic Universe*, World Scientific Publishing Co., River Edge, New Jersey

Guzzo M. and Morbidelli A. (1997) Construction of a Nekhoroshev-like result for the asteroid belt dynamical system, *Celest. Mech.*, **66**, 255.

Hartmann W.K. and Davis D.R. (1975) Satellite-sized planetesimals and lunar origin. *Icarus*, **24**, 504.

Hénon M. (1974) Integrals of the Toda lattice. *Phys. Rev. B*, **9**, 1921.

Hénon M. and Heiles C. (1964) The applicability of the third integral of motion: some numerical experiments. *Astron. J.*, **69**, 73.

Henrard J. (1970) On a perturbation theory using Lie transforms. *Celest. Mech.*, **3**, 107.

Henrard J. (1990) A semi-numerical perturbation method for separable Hamiltonian systems. *Celest. Mech.*, **49**, 43.

Henrard J. and Caranicolas N.D. (1990) Motion near the 3/1 resonance of the planar elliptic restricted three body problem. *Celest. Mech.*, **47**, 99.

Henrard J. and Henrard M. (1991) Slow crossing of a stochastic layer. *Physica D*, **54**, 135.

Henrard J. and Lemaître A. (1983) A second fundamental model for resonance. *Celest. Mech.*, **30**, 197.

Henrard J. and Lemaître A. (1987) A perturbative treatment of the 2/1 Jovian resonance. *Icarus*, **69**, 266.

Henrard J., Watanabe N. and Moons M. (1995) A bridge between secondary and secular resonances inside the Hecuba gap. *Icarus*, **115**, 336.

Hirayama K. (1918) Groups of asteroids probably of common origin. *Astron. J.*, **31**, 185.

Holman M.J. and Murray N.W. (1996) Chaos in high order mean motion resonances in the outer asteroid belt. *Astron. J.*, **112**, 1278.

Holman M.J. and Wisdom J. (1993) Dynamical stability in the outer solar system and the delivery of short period comets. *Astron. J.*, **105**, 1987.

Hori G.I. (1966) General perturbations theory with unspecified canonical variables. *Publ. Astron. Soc. Japan*, **18**, 287.

Ida S., Canup R.M. and Stewart G.R. (1997) Lunar accretion from an impact-generated disk. *Nature*, **389**, 353.

Jefferys W.H. and Standish E.M. (1972) Further periodic solutions of the three dimensional restricted problem. *Astron. J.*, **77**, 394.

Jewitt D.C., Luu J.X. and Trujillo C. (1998) Large Kuiper belt objects: the Mauna Kea 8k CCD survey. *Astron. J.*, **115**. 2125.

Kaneko K. and Konishi T. (1994) Peeling the onion of order and chaos in a high dimensional Hamiltonian system. *Physica D*, **71**, 146.

Kinoshita H. and Nakai H. (1996) Long-term behavior of the motion of Pluto over 5.5 billion years. *Earth Moon and Pl.*, **72**, 165.

Kinoshita H. and Nakai H. (1999) Analytical solution of the Kozai resonance and its application. *Celest. Mech.*, **75**, 125.

Kinoshita H. and Souchay J. (1990) The theory of the nutation for the rigid Earth model at the second order. *Celest. Mech.*, **48**, 187.

Kirkwood D. (1866). In *Proceedings of the American Association for the Advancement of Science for 1866.*

Knežević Z., Milani A., Farinella P., Froeschlé Ch. and Froeschlé C. (1991) Secular resonances from 2 to 50 AU. *Icarus*, **93**, 316.

Knežević Z., Froeschlé Ch., Lemaître A., Milani A. and Morbidelli A. (1995) Comparison between two thories of asteroid proper elements. *Astron. Astrophys.*, **293**, 605.

Kolmogorov A.N. (1954) Preservation of conditionally periodic movements with small change in the Hamiltonian function. *Dokl. Akad. Nauk SSSR*, **98**, 527; english translation in *Lecture notes in Physics*, **93** Springer-Verlag, Berlin, 51.

Kolmogorov A.N. (1959) On the entropy per time unit as a metric invariant of automorphisms. *Dokl. Acad. Nauk. SSSR*, **124**, 754.

Koseleff P.V. (1993) Relations among Lie formal series and construction of symplectic integrators. In *Applied Algebra, Algebraic Algorithms and Error Correcting Codes*, *Lecture Notes in Comp. Sci.*, **673**, Springer-Verlag, Berlin, 213.

Koseleff P.V. (1996) Exhaustive search of symplectic integrators using computer algebra. In *Integration Algorithms and Classical Mechanics*, *Fields Inst. Commun.*, **10**, Amer. Math. Soc., Providence, R.I., 103.

Kozai Y. (1962) Secular perturbations of asteroids with high inclination and eccentricities. *Astron. J.*, **67**, 591.

Kozai Y. (1978) Secular perturbations of asteroids and comets. In *Dynamics of the Solar System.* Reidel, Dordrecht, 231.

Krylov N. (1950) Studies on the foundation of statistical physics. *Akad. Nauk. SSSR*, M.-L.

Laskar J. (1985) Accurate methods in general planetary theory. *Astron. Astrophys.*, **144**, 133.

Laskar J. (1986) Secular terms of classical planetary theories using the results of general theory. *Astron. Astrophys.*, **157**, 59.

Laskar J. (1988) Secular evolution of the solar system over 10 million years. *Astron. Astrophys.*, **198**, 341.

Laskar J. (1989) A numerical experiment on the chaotic behaviour of the solar system. *Nature*, **338**, 237.

Laskar J. (1990) The chaotic motion of the solar system: a numerical estimate of the size of the chaotic zones. *Icarus*, **88**, 266.

Laskar J. (1992) A few points on the stability of the Solar System. In *Chaos, Resonance and Collective Dynamical Phenomena in the Solar System*, S. Ferraz-Mello, ed. Kluwer, Dordrecht, 1.

Laskar J. (1993) Frequency analysis for multidimensional systems. *Phys. D*, **67**, 257.

Laskar J. (1994) Large-scale chaos in the solar system. *Astron. Astrophys.*, **287**, 9.

Laskar J. (1999) Introduction to frequency map analysis. in *Hamiltonian Systems with Three or More Degrees of Freedom*, C. Simó, ed. Kluwer, Dordrecht, 134.

Laskar J. and Robutel Ph. (1993) The chaotic obliquity of the planets. *Nature*, **361**, 608.

Laskar J. and Robutel Ph. (1995) Stability of the planetary three-body problem. I. Expansion of the planetary Hamiltonian. *Celest. Mech.* **62**, 193.

Laskar J., Froeschlé C. and Celletti A. (1992a) The measure of chaos by the numerical analysis of the fundamental frequencies. *Phys. D*, **56**, 253.

Laskar J., Quinn T. and Tremaine S. (1992b) Confirmation of resonant structure in the solar system. *Icarus*, **95**, 148.

Laskar J., Joutel F. and Robutel Ph. (1993) Stabilization of the Earth's obliquity by the Moon. *Nature*, **361**, 615.

Lazutkin V.F. (1973) The existence of caustics for a billiard problem in a convex domain. *Math. USSR Izis*, **7**, 185.

Lemaître A. (1984) High-order resonances in the restricted three-body problem. *Celest. Mech.*, **32**, 109.

Lemaître A. and Henrard J. (1990) On the origin of the chaotic behavior in the 2/1 Kirkwood gap. *Icarus*, **83**, 391.

Lemaître A. and Morbidelli A. (1994) Calculation of proper elements for high inclined asteroidal orbits. *Celest. Mech.*, **60**, 29.

Levison H.F. and Duncan M.J. (1994) The long-term dynamical behavior of short-period comets, *Icarus*, **108**, 18.

Lichtenberg A.J. and Lieberman M.A. (1983) *Regular and Stochastic Motion*, Springer-Verlag, New York.

Lochak P. (1992) Canonical perturbation theory via simultaneous approximations. *Russ. Math. Surv.*, **47**, 57.

Lyapunov A.M. (1907) Problème général de la stabilité du mouvement. *Ann. Fac. Sci. Univ. Toulouse*, **9**, 203. Reproduced in *Ann. Math. Study*, **17**, Princeton (1947).

Malhotra R. (1996) The phase space structure near Neptune resonances in the Kuiper belt. *Astron. J.*, **111**, 504.

Marzari F., Davis D. and Vanzani V. (1995) Collisional evolution of asteroid families. *Icarus*, **113**, 168.

Marzari F., Farinella P. and Davis D.R. (1999) Origin, aging, and death of asteroid families. *Icarus*, **142**, 63.

Matese J.J. and Whitman P.G. (1992) A model of the galactic tidal interaction with the Oort comet cloud. *Celest. Mech.*, **54**, 13.

Melnikov V.K. (1963) On the stability of the center for time periodic perturbations. *Trans. Moscow Math. Soc.*, **12**, 1.

Menichella M., Paolicchi P. and Farinella P. (1996) The main belt as a source of Near Earth Asteroids. *Earth, Moon and Planets*, **72**, 133.

Message J. (1958) The search for asymmetric periodic orbits in the restricted problem of three bodies. *Astron. J.*, **63**, 443.

Message P.J. (1966) On nearly-commensurable periods in the restricted problem of three bodies, with calculation of the long-period variations in the interior 2/1 case. In *The Theory of Orbits in the Solar System and in Stellar Systems*, G. Contopoulos, ed. Academic Press, London, 197.

Michel P. (1997) Effects of linear secular resonances in the region of semimajor axes smaller than 2 AU. *Icarus*, **129**, 348.

Michel P. and Froeschlé Ch. (1997) The location of linear secular resonances for semimajor axes smaller than 2 AU. *Icarus*, **128**, 230.

Michel P. and Thomas F. (1996) The Kozai resonance for near-Earth asteroids with semimajor axes smaller than 2 AU. *Astron. Astrophys.*, **307**, 310.

Michel P., Migliorini F., Morbidelli A. and Zappalà V. (2000) The population of Mars-Crossers: classification and dynamical evolution. *Icarus*, **145**, 332.

Michtchenko T.A. and Ferraz-Mello S. (1995) Comparative study of the asteroidal motion in the 3:2 and 2:1 resonances with Jupiter. I: planar model. *Astron. Astrophys.*, **303**, 945.

Michtchenko T.A. and Ferraz-Mello S. (1997) Escape of asteroids from the Hecuba gap. *Planet. Sp. Sci.*, **45**, 1587.

Migliorini F., Michel P., Morbidelli A., Nesvorný D. and Zappalà V. (1998) Origin of Earth-crossing asteroids: a quantitative simulation, *Science*, **281**, 2022.

Milani A. and Knežević Z. (1990) Secular perturbation theory and computation of asteroid proper elements. *Celest. Mech.*, **49**, 247.

Milani A. and Knežević Z. (1992) Asteroid proper elements and secular resonances. *Icarus*, **98**, 211.

Milani A. and Knežević Z. (1994) Asteroid proper elements and the dynamical structure of the asteroid main belt. *Icarus*, **107**, 219.

Milani A., Nobili A.M. and Carpino M. (1989) Dynamics of Pluto. *Icarus*, **82**, 200.

Moons M. (1993) Averaging approaches. *Dep. of Math. Rep.*, **19**, FUNDP, Namur, Belgium.

Moons M. (1994) Extended Schubart averaging. *Celest. Mech.*, **60**, 173.

Moons M. (1997) Review of the dynamics in the Kirkwood gaps. *Celest. Mech.*, **65**, 175.

Moons M. and Morbidelli A. (1993) The main mean motion commensurabilities in the planar circular and elliptic problem. *Celest. Mech.*, **57**, 99.

Moons M. and Morbidelli A. (1995) Secular resonances inside mean-motion commensurabilities: the 4/1, 3/1, 5/2 and 7/3 cases. *Icarus*, **114**, 33.

Moons M., Morbidelli A. and Migliorini F. (1998) Dynamical structure of the 2/1 commensurability and the origin of the resonant asteroids. *Icarus*, **135**, 458.

Morbidelli A. (1993a) Asteroid secular resonant proper elements. *Icarus*, **105**, 48.

Morbidelli A. (1993b) On the successive elimination of perturbation harmonics. *Celest. Mech.*, **55**, 101.

Morbidelli A. (1997) Chaotic diffusion and the origin of comets from the 2/3 resonance in the Kuiper belt. *Icarus*, **127**, 1.

Morbidelli A. and Giorgilli A. (1993) Quantitative perturbation theory by successive elimination of harmonics. *Celest. Mech.*, **55**, 131.

Morbidelli A. and Giorgilli A. (1995a) Superexponental stability of KAM tori. *Journal Stat. Phys.*, **78**, 1607.

Morbidelli A. and Giorgilli A. (1995b) On a connection between KAM and Nekhoroshev theorems. *Physica D*, **86**, 514.

Morbidelli A. and Giorgilli A. (1997) On the role of high order resonances in normal forms and in separatrix splitting. *Physica D*, **102**, 195.

Morbidelli A. and Guzzo M. (1996) The Nekhoroshev theorem and the asteroid belt dynamical system. *Celest. Mech.*, **65**, 107.

Morbidelli A. and Henrard J. (1991a) Secular resonances in the asteroid belt: theoretical perturbation approach and the problem of their location. *Celest. Mech.*, **51**, 131.

Morbidelli A. and Henrard J. (1991b) The main secular resonances ν_6, ν_5 and ν_{16} in the asteroid belt. *Celest. Mech.*, **51**, 169.

Morbidelli A. and Moons M. (1993) Secular resonances inside mean motion commensurabilities: the 2/1 and 3/2 cases. *Icarus*, **103**, 99.

Morbidelli A. and Nesvorný D. (1999) Numerous weak resonances drive asteroids towards terrestrial planets orbits. *Icarus*, **139**, 295.

Morbidelli A., Thomas F. and Moons M. (1995a) The resonant structure of the Kuiper belt and the dynamics of the first five trans-Neptunian objects. *Icarus*, **118**, 322.

Morbidelli A., Zappalà V., Moons M., Cellino A. and Gonczi R. (1995b) Asteroid families close to mean motion resonances: dynamical effects and physical implications. *Icarus*, **118**, 132.

Moser J. (1962) On invariant curves of area-preserving mappings of an anulus. *Nachr. Akad. Wiss. Gottingen, Math. Phys.* **2**, 1.

Murray N. and Holman M. (1997) Diffusive chaos in the outer asteroid belt. *Astron. J.*, **114**, 1246.

Murray N. and Holman M. (1999) The origin of chaos in the outer solar system. *Science*, **283**, 1877.

Murray N., Holman M. and Potter M. (1998) On the origin of chaos in the asteroid belt. *Astron. J.*, **116**, 2583.

Nakai H. and Kinoshita H. (1985) Secular perturbations of asteroids in secular resonances. *Celest. Mech.*, **36**, 391.

Neishtadt A.I. (1982) Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions. *PMM USSR*, **45**, 766.

Neishtadt A.I. (1984) The separation of motions in systems with rapidly rotating phase. *J. Appl. Math. Mech.*, **48**, 133.

Nekhoroshev N.N. (1977) Exponential estimates of the stability time of near-integrable Hamiltonian systems. *Russ. Math. Surveys*, **32**, 1.

Nekhoroshev N.N. (1979) Exponential estimates of the stability time of near-integrable Hamiltonian systems, II. *Trudy Sem. Petrovs.*, **5**, 5.

Néron de Surgy O. and Laskar, J. (1997) On the long term evolution of the spin of the Earth. *Astron. Astrophys*, **318**, 975.

Nesvorný D. (2001) How long-lived are the hypothetical Trojan populations of Saturn, Uranus and Neptune?, *Icarus*, submitted.

Nesvorný D. and Ferraz-Mello S. (1997a) Chaotic diffusion in the 2/1 asteroidal resonance: an application of the frequency map analysis. *Astron. Astrophys.*, **320**, 672.

Nesvorný D. and Ferraz-Mello S. (1997b) On the asteroidal population of the first-order Jovian resonances. *Icarus*, **130**, 247.

Nesvorný D. and Morbidelli A. (1998) Three-body mean motion resonances and the chaotic structure of the asteroid belt, *Astron. J.*, **116**, 3029.

Nesvorný D. and Morbidelli A. (1999) An analytic model of three-body mean motion resonances. *Celest. Mech.*, **71**, 243.

Nesvorný D. and Roig F. (2000) Mean motion resonances in the trans-neptunian region: Part I: The 2:3 resonance with Neptune. *Icarus*, in press.

Nesvorný D. and Roig F. (2001) Mean motion resonances in the trans-neptunian region: Part II: the 1:2, 3:4 and weaker resonances. *Icarus*, in press.

Nesvorný D., Roig F. and Ferraz-Mello S. (2000) Close approaches of trans-Neptunian objects to Pluto have left observable signatures on their orbital distribution. *Astron. J.*, **119**, 953.

Nobili A.M., Milani A. and Carpino M. (1989) Fundamental frequencies and small divisors in the orbits of the outer planets *Astron. Astrophys.*, **210**, 313.

Olvera A. and Simó C. (1987) An obstruction method for the destruction of invariant curves. *Physica D*, **26**, 181.

Öpik E.J. (1976) *Interplanetary Encounters.* Elsevier, New York.

Oseledec V.I. (1968) The multiplicative ergodic theorem: the Laypunov characteristic numbers of dynamical systems. *Trudy Mosk. Mat. Obsch.*, **19**, 179. English translation in *Trans. Mosc. Soc.*, **19**, 197 (1968).

Papaphilippou Y. and Laskar J. (1998) Global dynamics of triaxial galactic models through frequency map analysis. *Astron. Astropys.*, **329**, 451.

Parisi M.G. and Brunini A. (1996) Dynamical consequencies of Uranus' great collision. In *Chaos in gravitational N-body systems*, J.C. Muzzio, S. Ferraz-Mello and J. Henrard, eds. Kluwer, Dordrecht, 291.

Pesin Y.B. (1977) Lyapunov characteristic exponents and smooth ergodic theory. *Russ. Math. Surveys*, **32**, 55.

Piessens R., deDoncker-Kapenga E., Uberhuber C. W., and Kahaner D. K. (1983) *Quadpack.* Springer-Verlag, New–York.

Poincaré H. (1892) *Les méthodes nouvelles de la mécanique céleste.* Gauthier-Villar, Paris.

Poincaré H. (1902a) Les solutions périodiques et les planètes du type d'Hécube. *Bull. Astron.*, **19**, 177.

Poincaré H. (1902b) Sur les planètes du type d'Hécube. *Bull. Astron.*, **19**, 289.

Pöschel J. (1993) Nekhoroshev's estimates for quasi-convex Hamiltonian systems. *Math. Z.*, **213**, 187.

Press W.H., Teukolsky S.A., Vetterling W.T. and Flannery B.P. (1986) *Numerical Recipes.* Cambridge University Press.

Quinn T.R., Tremaine S. and Duncan M.J. (1991) A three million year integration of the Earth's orbit. *Astron. J.*, **101**, 2287.

Robutel Ph. and Laskar J. (2001) Frequency map and global dynamics in the solar system I: short periodic dyanmics of massless particles. *Icarus*, in press.

Safronov V.S. (1966) Sizes of the largest bodies falling onto the planets. *Sov. Astron.*, **9**, 987.

Scholl H. and Froeschlé Ch. (1990) Orbital evolution of known asteroids in the ν_5 resonance region. *Astron. Astrophys.*, **227**, 255.

Schubart J. (1964) Long period effects in nearly commensurable cases of the restricted three-body problem. *Smithsonian Astrophys. Obs. Spec. Rep.*, **149**.

Schubart J. (1968) Long-period effects in the motion of Hilda-type planets. *Astron. J.*, **73**, 99.

Sessin W. and Ferraz-Mello S. (1984) Motion of two planets with periods commensurable in the ratio 2:1 solutions of the Hori auxiliary system. *Celest. Mech.*, **32**, 307.

Šidlichovský M. (1990) *Tables of the Disturbing Function for Resonant Asteroids.* Czechoslovak Academy of Sciences, Prague.

Simó C. (1990) On the analytical and numerical approximation of invariant manifolds. In *Modern Methods in Celestial Mechanics*, D. Benest and C. Froeschlé, eds. Ed. Frontières, Gif sur Yvettes, France.

Simon J.L. and Bretagnon P. (1975) First order perturbations of the four big planets: literal variations. *Astron. and Astrophys.*, **42**, 259.

Sinai Ya.G. (1959) On the concept of entropy of a dynamical system. *Dokl. Acad. Nauk. SSSR*, **124**, 768.

Smale S. (1963) Diffeomorphism with many periodic points. In *Differential and Combinatorial Topology*, S.S. Cairns, ed. Princeton University Press, 63.

Smale S. (1980) *The Mathematics of Time: Essays on Dynamical Systems, Economic Processes and Related Topics.* Springer-Verlag, New York.

Sommerfeld A. (1922) *Atombau und Spektrallinien.* Vieweg, Braunschweig, Germany.

Souchay J., Loysel B., Kinoshita H. and Folgueira M. (1999) Corrections and new developments in rigid earth nutation theory. III. Final tables "REN-2000" including crossed-nutation and spin-orbit coupling effects. *Astron. Astrophys. Suppl. Ser.*, **135**, 111.

Sussman G.J. and Wisdom J. (1988) Numerical evidence that the motion of Pluto is chaotic. *Science*, **241**, 433.

Sussman G.J. and Wisdom J. (1992) Chaotic evolution of the solar system. *Science*, **257**, 56.

Taylor S.R. (1999) On the difficulties of making Earth-like planets. *Meteoritics. Pl. Sci.*, **34**, 317.

Thomas F. (1998) 'La dynamique résonnante dans le système solaire. Application au mouvement des objets transneptuniens'. Ph.D. thesis, Obs. de Paris.

Thomas F. and Morbidelli A. (1996) The Kozai resonance in the outer Solar System and the dynamics of long-period comets, *Celest. Mech.*, **64**, 209.

Touma J. and Wisdom J. (1993) The chaotic obliquity of Mars. *Science*, **259**, 1294.

Touma J. and Wisdom J. (1994a) Evolution of the Earth–Moon system. *Astron. J.*, **108**, 1943.

Touma J. and Wisdom J. (1994b) Lie–Poisson integrators for rigid body dynamics in the Solar System. *Astron. J.*, **107**, 1189.

Ward W.R. (1974) Climatic variations on Mars: I. Astronomical theory of insolation. *J. Geophys. Res.*, **79**, 3375.

Ward W.R. (1982) Comments on the long-term stability of the Earth's obliquity. *Icarus*, **50**, 444.

Wetherill G.W. (1975) Late heavy bombardment of the Moon and terrestrial planets. *Proc. Lunar Sci. Conf.*, **6**, 1539.

Whittaker E.T. (1937) *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies.* Cambridge University Press.

Wiggins S. (1988) *Global Bifurcations and Chaos.* Springer-Verlag, New York.

Williams J.G. (1969) 'Secular perturbations in the solar system'. Ph.D. dissertation, University of California, Los Angeles.

Williams J.G. and Benson G.S. (1971) Resonances in the Neptune–Pluto system. *Astron. J.*, **76**, 167.

Williams J.G. and Faulkner J. (1981) The position of secular resonance surfaces. *Icarus*, **46**, 390.

Wisdom J. (1980) The resonance overlap criterion and the onset of stochastic behavior in the restricted three-body problem. *Astron. J.*, **85**, 1122.

Wisdom J. (1982) The origin of the Kirkwood gaps: a mapping for asteroidal motion near the 3/1 commensurability. *Astron. J.*, **87**, 577.

Wisdom J. (1983) Chaotic behavior and the origin of the 3/1 Kirkwood gap. *Icarus*, **56**, 51.

Wisdom J. (1985) A perturbative treatment of motion near the 3/1 commensurability. *Icarus*, **63**, 272.

Wisdom J. (1987) Urey Prize Lecture: chaotic dynamics in the solar system. *Icarus*, **72**, 241.

Wisdom J. and Holman M. (1991) Symplectic maps for the N-body problem. *Astron. J.*, **102**, 1528.

Yoshikawa M. (1987) A simple analytical model for the ν_6 resonance. *Celest. Mech.*, **40**, 233.

Yoshikawa M. (1990) Motions of asteroids at the Kirkwood gaps. I: on the 3:1 resonance with Jupiter. *Icarus*, **87**, 78.

Yoshikawa M. (1991) Motions of asteroids at the Kirkwood Gaps. II: on the 5:2, 7:3, and 2:1 resonances with Jupiter. *Icarus*, **92**, 94.

Yuasa M. (1973) Theory of secular perturbations of asteroids including terms of higher order and higher degree. *Publ. Astron. Soc. Japan*, **25**, 399.

Zappalà V., Cellino A., Farinella P. and Knežević Z. (1990) Asteroid families. I: identification by hierarchical clustering and reliability assessment. *Astron. J.*, **100**, 2030.

Zappalà V., Bendjoya Ph., Cellino A., Farinella P. and Froeschlé C. (1995) Asteroid families: Search of a 12,487-asteroid sample using two different clustering techniques. *Icarus*, **116**, 291.

Ziglin S.L. (1980) Non-integrability of a problem on the motion of four point vortices. *Sov. Math. Dokl.*, **21**, 296.

Index

N-body problem, *see* planetary problem

action–angle variables, 28–29
  resonant, 74–77
Amor asteroids, 4
angle
  helicity, *see* helicity angle
  of great inequality, *see* great inequality
  resonant, *see* resonance, critical angle
  twist, *see* twist angle
anomaly
  eccentric, *see* eccentric anomaly
  mean, *see* mean anomaly
  true, *see* true anomaly
aphelion, 13
  distance, 13
apocenter, 13
apocentric librator, 201
apogee, 13
Apollo asteroids, 4
argument of perihelion, 15
Arnold
  diffusion, 118
  web, 114
Arnold–Liouville theorem, 28–29, 74, 166
asteroid belt, *see* main asteroid belt
asteroid family, 4, 172–174, 179, 229, 291
Aten asteroids, 4
attractor, 25
averaging, *see also* Birkhoff normal form, 51–57

Bessel functions, 217, 237
Birkhoff normal form, 45, 47

canonical transformation, 22–24
Cassini
  division, 8
  states, 142
chaos, 1, 27, 81–114, 121–122
  in planet motion, 134–147, 254–256
  in small body motion, 210–229, 240–254, 262–266, 274, 280–330
chaotic
  diffusion, *see* diffusion
  region, 70, 81, 84
Chirikov
  criterion, 109
  diffusion, 111, 122
circulation, 73
close encounter, 138, 159, 283, 285, 315, 317
  protection from, 206–207, 292, 319
collision
  curve, 206, 210
  with the Sun, 165, 187, 285
Colombo's top, 142
comet, 6–8, 162–165

Halley-type, 7
Jupiter-family, 7
long-period, 7, 162–165
Sun-grazer, 165
commensurability, *see* mean motion resonance
constant of motion, 25, 27, 28
of the Keplerian problem, 30, 38
convexity, 115, 118, 188
coordinates, 18
barycentric, 11
heliocentric, 11, 17, 22–24
rotating, *see* rotating frame
spherical, 19

D'Alembert rules, 35–36
degeneracy, 37–38, 53, 159, 195, 222
degrees of freedom, 18
Delaunay variables, 29–35
mean, 54, 127, 149
modified, 35
semimean, 57, 196, 233
diffusion, 66, 111, 312–327
Arnold, *see* Arnold, diffusion
Chririkov, *see* Chirikov, diffusion
coefficient, 323–327
diophantine condition, 44
Dirac delta function, 322
domain
Nekhoroshev, *see* Nekhoroshev, domain
nonresonant, 44, 117, 119, 122, 126
resonant, *see* resonance, domain

Earth, 138, 142–147
eccentric anomaly, 14
eccentricity, 13
Edgeworth–Kuiper belt, *see* Kuiper belt
elliptic function, 76, 156
Encke division, 8
equation
Fokker–Planck, *see* Fokker–Planck equation
homologic, *see* homologic equation
Kepler, *see* Kepler, equation
Lagrange, *see* Lagrange equations
equilibrium point
elliptic, *see* equilibrium point, stable
hyperbolic, *see* equilibrium point, unstable
stable, 72, 73
unstable, 72, 73, 81
equinox precession, 141
expansion
Fourier, *see* Fourier expansion
Legendre, *see* Legendre expansion
Taylor, *see* Taylor expansion

fast drift direction, 118
Fokker–Planck equation, 323
Fourier expansion, 43
frequency, 37
analysis, 97–103
completely resonant, 37
diophantine, 44, 60, 63
nonresonant, 37
proper, *see* proper frequency
space, 66, 117
function
Bessel, *see* Bessel functions
Dirac, *see* Dirac delta function
disturbing, *see* perturbation
elliptic, *see* elliptic function
generating, 24
Hamiltonian, *see* Hamiltonian
time evolution of, 25–26

Galactic tide, 8, 160

great inequality, 278, 295
Greek asteroid, 4, 209

Hamiltonian, 18
  flow properties, 24–26
  generating, 24, 41–45
  integrable, 26–29
  nonautonomous, 21
  of the N-body problem, *see* Hamiltonian, of the planetary problem
  of the planetary problem, 21–24, 52
  of the restricted problem, 21, 52
  quasi-integrable, 39
  time-dependent, *see* Hamiltonian, nonautonomous
harmonic
  generation of, 49–51, 232–239
  order of, 43
  successive elimination of, 257–260
helicity angle, 104
heteroclinic
  intersection, *see* heteroclinic, point
  point, 111–114
Hildas, 292–293
Hill's radius, 210
homoclinic
  intersection, *see* homoclinic, point
  point, 83, 84, 86
  tangle, 84
homologic equation, 41
horseshoe
  librator, 209
  of Smale, *see* Smale horseshoe
hyperbolic point, *see* equilibrium point, unstable
hyperbolicity, 81, 94

inclination, 14
integrability, 27
invariant plane, 131
island, 70, 80, 162, 164, 165, 214, 224
isochronous system, 38

Jupiter, 1, 217, 254–256

KAM
  theorem, 59–63, 327–328
  torus, *see* torus, KAM
Kepler
  equation, 14
  laws, 12
  problem, 11–16, 29
Kirkwood gap, 3, 8, 214, 229, 280, 285, 286
Kolmogorov
  normal form, 62
  theorem, *see* KAM, theorem
Kozai
  dynamics, *see* Kozai, resonance
  Hamiltonian, 155, 268
  resonance, 154–167, 172, 177, 179, 185–186
    inside mean motion resonance, 266–271, 285, 288, 292, 297, 299
KS–entropy, 92, 94
Kuiper belt, 5–7, 179, 187, 214–216, 253–254, 294–300, 310, 312, 317–320, 329

Lagrange equations, 17, 34
Lagrange–Laplace solution, 130
  coefficients of, 133–135
Lagrangian points, 209
Laplace
  resonance, 9, 217, 231
  solution, *see* Lagrange–Laplace solution

Legendre expansion, 235
libration
  region, 73, 199
  torus, *see* torus, librational
librator
  apocentric, *see* apocentric librator
  asymmetric, 202, 205, 206, 209, 298
  horseshoe, *see* horseshoe, librator
  tadpole, *see* tadpole librator
Lie series, 26, 40–42
Liouville's theorem, 25, 27
longitude of node, 15
longitude of perihelion, 16
Lyapunov
  exponent, 92–97
    fast indicator, 103–104
  orbit, 210

main asteroid belt, 2–4, 156–160, 170–179, 183–193, 212–214, 247–253, 280–294, 303–308, 313–317, 323, 325–327, 329
manifold
  stable, 82–86
  unstable, 82–86
Mars, 138, 147
mean anomaly, 14
mean elements, 54
mean longitude, 16
mean motion, 14
mean motion resonance, 3, 6, 8, 9, 54, 55, 195–229, 257–330
  1/1, 207–210, 218–220, 300–302
  1/2 with Neptune, 297–300
  2/1 with Jupiter, 286–292
  2/3 with Neptune, 295–297
  3/1 with Jupiter, 280–286
  3/2 with Jupiter, 292–294
  critical angle, 197
  eccentricity order, 197
  exterior, 199
  interior, 199
  modulation of, 224–229, 272, 281, 308
  multiplet, *see* multiplet
  normal form, 57, 196, 260
  of 1st eccentricity order, 199–202, 286–300
  of 2nd eccentricity order, 205, 280–286
  of 3rd eccentricity order, 206
  of 4th eccentricity order, 206
  overlapping, 210–220
  with Mars, 306–308, 315
mean variables, *see* Delaunay variables, mean
MEGNO, 104–106
Melnikov integral, *see* Poincaré–Melnikov integral
Mercury, 138, 140, 147
module of the resonance, *see* resonance, module
momenta, 18
Moon formation, 140, 146
motion
  chaotic, *see* chaos
  constant of, *see* constant of motion
  Keplerian, *see* Kepler, problem
  periodic, 37
  quasi-periodic, 37
multiplet, 220–224, 231, 240–246, 305, 320–322, 326
multiplicity, *see* resonance, multiplicity of

near-Earth asteroids, 4–5, 177, 187, 285, 317
Nekhoroshev
  domain, 328

structure, 119–124, 126, 328–330
system, 122
theorem, 114–119, 327–330
Neptune, 217, 254–256
nodal distance, 157
node, 14
longitude, *see* longitude of node
node crossing curve, 159–165
normal form
Birkhoff, *see* Birkhoff normal form
Kolmogorov, *see* Kolmogorov, normal form
mean motion resonance, *see* mean motion resonance, normal form
optimal, 48–49
remainder of, 48, 78–79
resonant, *see* resonance, normal form
secular, *see* secular normal form
secular resonant, *see* secular resonance, normal form
three-body resonance, *see* three-body resonance, normal form

obliquity, 141
Oort cloud, 7, 160
orbital elements
mean, *see* mean elements
osculating, *see* osculating elements
proper, *see* proper elements
semimean, *see* semimean elements
semiproper, *see* semiproper elements
order
in eccentricity, *see* mean motion resonance, eccentricity order
of the harmonic, *see* harmonic, order of
of the resonance, *see* resonance, order of
osculating elements, 12–16, 18
formal, 18, 34

pendulum, 72–74, 82
modulated, 223–224, 326–327
pericenter, 13
perigee, 13
perihelion, 13
argument, *see* argument of perihelion
distance, 13
longitude, *see* longitude of perihelion
perturbation, 16–18, 20–22, 39
perturbation theory, 39–49
phase space, 18
planetary problem, 16–17, 22–24
Plutinos, 216, 295, 297
Pluto, 5, 6, 206, 216, 271, 295, 297
Poincaré section, 67–68, 80, 89, 218, 224, 243, 264, 274
Poincaré variables, 36
Poincaré–Melnikov integral, 84–86
Poisson bracket, 22
practical stability, 114, 115, 179, 256
proper elements, 4, 167–174, 250, 313–315
linear, 154
proper frequency, 131, 169
linear, 154
proper variables, 169
linear, 154

reduced mass, 20, 23
remainder, *see* normal form, remainder of
resonance

action–angle variables, *see* action–angle variables, resonant
critical angle, 72
of mean motion resonance, *see* mean motion resonance, critical angle
of secular resonance, *see* secular resonance, critical angle
definition of, 37
domain, 118–119, 122
interactions, 107–126
island, *see* island
Kozai, *see* Kozai, resonance
mean motion, *see* mean motion resonance
module, 47
multiplet, *see* multiplet
multiplicity of, 37
normal form, 44–48, 71
of Laplace, *see* Laplace, resonance
order of, 37
overlapping, 108–111, 114, 122–126, 145–146, *see also* mean motion resonance, overlapping, *see also* multiplet
region, 68, 73, 78, 199, 214
secondary, 78, 262–266
secular, *see* secular resonance
separatrix, *see* separatrix
single resonance, 71–88, 107, 118, 122
three-body, *see* three-body resonance
torus, *see* torus, resonant
web, *see* Arnold, web
restricted problem, 17, 21
planar circular, 196
dynamics of, 196–220
Ricci curvature, 92
ringlets, 9
rings, 8–9
rotating frame, 19–20

satellite
Galilean, *see* satellite, of Jupiter
of Jupiter, 9, 217
of Uranus, 160
prograde, 210
retrograde, 210
shepherding, 9
Saturn, 1, 217, 254–256
secular dynamics
of planets, 127–147
of small bodies, 149–193, 266–277
secular normal form, 53–57, 127, 149, 328
secular resonance, 3, 6, 152, 169, 174–193, 270–277, 310
ν_5, 177, 183–185, 188, 271, 272, 274, 280–283, 285, 288, 292
ν_6, 177, 182, 187–193, 271, 274, 283, 285, 288, 292, 300, 306, 328
ν_{16}, 177, 182, 185–186, 276, 285, 288–289, 292
ν_{17}, 179, 187, 310
ν_{18}, 179, 187, 276, 295, 297, 310
ν_7, 179, 187, 319
ν_8, 179, 187, 271, 295, 310, 319
critical angle, 181
in mean motion resonances, 270–277
Kozai, *see* Kozai, resonance
linear, 152, 174
normal form, 180–181, 184, 186–188, 190
of first order, 152, 180
semimajor axis, 13
semimean elements, 57

semimean variables, *see* Delaunay variables, semimean
semiproper elements, 182
semiproper variables, 180, 181
separatrix, 73, 77, 108
  splitting of, 81–88, 118
Smale horseshoe, 84
small divisors, 42–44
spin axis dynamics, 140–147
splitting, *see* separatrix, splitting of
standard map, 68–70, 79–81, 109, 113–114, 322

tadpole librator, 209
Taylor expansion, 26
three-body resonance, 231–256
  (5 −2 −2), 241–245
  (6 1 −3), 245–246
  in mean motion resonances, 277–280, 291, 293, 295
  in planet motion, 254–256
  multiplet, *see* multiplet
  normal form, 234, 240
torus, 37
  complex, 43
  invariant, 37, 59, 60, 63–67
  KAM, 59–70, 93, 98, 102, 106–108, 111, 121–122
    exponential stability, 67
    golden, 109
    isolating property, 64–66
    measure, 64
    persistence of, 63
    superexponential stability, 122
    translation of, 63
  librational, 79
  resonant, 79–81
transformation
  canonical, *see* canonical transformation
  close to the identity, 40
  polar to cartesian coordinates, 24, 36
Trojan asteroid, 4, 209, 219, 301
true anomaly, 14
twist angle, 104
two-body problem, *see* Kepler, problem

Uranus, 140, 147, 217, 254–256

variables
  action–angle, *see* action–angle variables
  conjugate, 18
  Delaunay, *see* Delaunay variables
  mean, *see* Delaunay variables, mean
  Poincaré, *see* Poincaré variables
  proper, *see* proper variables
  semimean, *see* Delaunay variables, semimean
  semiproper, *see* semiproper variables
Venus, 138, 147
volume conservation, 25

web, *see* Arnold, web

Yarkovsky effect, 286